Matrix Inverses: From Geometry To Permutations

Ross Ure Anderson *

22nd June, 2026

Abstract. This article gives a development from first principles of matrix inverses commencing with a geometric proof of the 3×3 case for $\mathbb{R}$ and progressing to the $n \times n$ case over a general field $\mathbb{F}$ using permutations. All the theory of permutations required is provided in the article [43]. No vector space theory is required apart from the definitions of vector space and subspace, span of a vector sequence, linear independence, and linear mapping. Only basic facts about matrices are assumed and proofs of all results on elementary row operations required for the $n \times n$ case are provided. Care is taken to distinguish between geometric arguments and 'a priori' algebraic arguments and a separate chapter develops the scalar product, vector product, scalar triple product, vector triple product, and right/left handedness properties in a purely geometric fashion without the aid of coordinate systems. Coordinate systems are then introduced and the problem of transforming coordinates to a general 'skew' coordinate system is considered which leads to a solution to the 3×3 matrix inverse over $\mathbb{R}$. Building on the insights from the real case we then consider the intermediate stage of a 3×3 matrix over a general field $\mathbb{F}$ and develop a set of properties of scalar and vector product on $\mathbb{F}^3$ sufficient to facilitate a purely algebraic proof of the 3×3 inverse over $\mathbb{F}$ which is independent of the earlier geometric proof for $\mathbb{R}$. The 3×3 inverse over $\mathbb{F}$ is then reinterpreted into the Leibniz and Laplace forms which introduces the connection to permutations. The generalization of these forms to the case of the $n \times n$ inverse is then conjectured and proved. Care is taken to distinguish between working in a general field $\mathbb{F}$ and in $\mathbb{R}$, in particular the case of the field GF(2) of characteristic 2 where division by 2 is not possible. Since $\mathbb{F}$ is a completely general field all the proofs given for $\mathbb{F}$ cover $n \times n$ matrices over finite fields such as integers mod p, which have many applications in areas such as cryptography and coding theory. A critical part of the process in the $n \times n$ case is equating the Leibniz form to the Laplace form (Theorem 32) and this is achieved using a theorem on permutation extensions provided in [43, §12]. A proof of the Cauchy-Binet formula is also given and the theorems on permutations it relies upon can be found in [43]. An emphasis is placed on the value of the geometry underlying matrix inverses since it provides a natural solution to the initial 3×3 case over $\mathbb{R}$ which leads to the initial form of the inverse that is the basis for the general $n \times n$ case over $\mathbb{F}$. An initial 'scholium' of a kind is provided on Euclidean space to review some core geometric concepts in a coordinate system free way and to link these concepts to 'a priori' mathematical arguments.

*E-mail: ruanderson100@yahoo.com, Web: https://archive.org/details/@ross_ure_anderson

Contents

1 Introduction **4**

2 Euclidean Space **5**

3 Euclidean Vectors **8**
3.1 Relationship Between EV And ES . . . 9
3.2 Scalar Product In EV . . . 10
3.2.1 EV As Direct Sum Of Line And Plane . . . 11
3.2.2 EV_2 As Direct Sum Of Two Lines . . . 12
3.3 Right-Hand Rule In EV . . . 14
3.4 Vector Product In EV . . . 17
3.5 Scalar Triple Product In EV . . . 19
3.6 Vector Triple Product In EV . . . 21

4 Coordinate Systems **22**
4.1 Cartesian Coordinate System . . . 22
4.1.1 EV Scalar Product Expressed In Cartesian Coordinates . . . 23
4.1.2 Determinant Notation . . . 24
4.1.3 Alternative Approach To Vector Product Using Scalar Product And CCS . . . 24
4.1.4 EV Vector Product Expressed In Cartesian Coordinates . . . 26
4.1.5 Scalar Triple Product Expressed In Cartesian Coordinates . . . 26
4.2 Scalar And Vector Product Operations For $\mathbb{R}^3$ . . . 27
4.3 Right And Left Handed Systems In $\mathbb{R}^3$ . . . 27
4.4 Skew Coordinate System . . . 28
4.5 Transformation Of Coordinates From Cartesian Coordinate System To Skew Coordinate System . . . 29
4.5.1 General 3D Rotation Of CCS . . . 30

5 Linear Independence **30**

6 Matrices **31**

7 Geometric Derivation Of Inverse Of 3 x 3 Matrix Over $\mathbb{R}$ **32**

8 Comments On The General Field $\mathbb{F}$ **33**

9 Vector Space $\mathbb{F}^3$ Over General Field $\mathbb{F}$ **35**
9.1 Definition Of Scalar And Vector Product On $\mathbb{F}^3$ . . . 35
9.2 Comparison Of $\mathbb{F}^3$ With Vector Spaces EV And $\mathbb{R}^3$ . . . 35
9.3 Properties Of Scalar And Vector Product On $\mathbb{F}^3$ . . . 36

10 Inverse Of 3 x 3 Matrix Over General Field $\mathbb{F}$ **38**
10.1 3 x 3 Product Formulae . . . 41

11 3 x 3 Matrix Inverse In Terms Of Minors, Cofactors, And Leibniz And Laplace Formulae **42**
11.1 3 x 3 Adjugate In Terms Of Minors And Cofactors . . . 42
11.2 3 x 3 Determinant In Leibniz Form . . . 43
11.3 3 x 3 Determinant In Laplace Form . . . 44

12 n x n Matrix Inverse **46**
12.1 n x n Determinant . . . 46
12.2 Determinant As A Multilinear Map . . . 48
12.3 n x n Minors And The Adjugate . . . 49
12.4 n x n Laplace Formulae . . . 50
12.4.1 Laplace Off-Row And Off-Column Expansions . . . 52
12.5 Elementary Row Operations . . . 53
12.6 Main Theorem . . . 56
12.7 n x n Product Formulae . . . 57

13 Cauchy–Binet Formula **58**
13.1 Proof Of Adjugate Product Formula . . . 61

Appendices **64**

Appendix A Permutations **64**

Appendix B Euler Angles **65**
B.1 Uniqueness Of Euler Angles 65
B.2 Elementary Rotation Matrices 66

Appendix C Completeness Of Real Numbers **67**

References **69**

1 Introduction

The following notations will be used :

- ES = Euclidean Space, EV = Euclidean Vectors
- DLS = directed line segment
- CS = coordinate system, CCS = Cartesian coordinate system
- $\mathbb{D}$ = the set of all directions in ES, $\mathbb{A}$ = the set of all planar angles
- $[n]$ = the set $\{1, 2, \ldots, n\}$
- $\mathbb{N}$ = the set of natural numbers $\{1, 2, 3, \ldots\}$
- $\mathbb{W}$ = the set of whole numbers (ie. non-negative integers), $\{0, 1, 2, 3, \ldots\}$
- $\mathbb{Z}$ = set of integers, $\mathbb{Q}$ = set of rational numbers, $\mathbb{R}$ = set of real numbers, $\mathbb{C}$ = set of complex numbers
- $\mathbb{F}$ = a general field
- $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ = plane contained by non-zero non-parallel vectors $\underline{\mathbf{a}}, \underline{\mathbf{b}} \in$ EV
- $\mathrm{RH}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ = right-handed side of plane $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$, $\mathrm{LH}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ = left-handed side of plane $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$
- $V(\underline{\mathbf{a}}, \underline{\mathbf{b}}, \underline{\mathbf{c}})$ = unsigned volume of parallelopiped defined by vectors $\underline{\mathbf{a}}, \underline{\mathbf{b}}, \underline{\mathbf{c}} \in$ ES
- $\underline{\mathbf{u}}$, $\underline{\mathbf{v}}$, $\underline{\mathbf{w}}$ = standard basis vectors in EV for a CCS
- $\underline{\mathbf{i}}$, $\underline{\mathbf{j}}$, $\underline{\mathbf{k}}$ = unit vectors (1, 0, 0), (0, 1, 0), and (0, 0, 1) in $\mathbb{R}^3$ or $\mathbb{F}^3$
- $\mathrm{sp}(\underline{\mathbf{v}}_1, \ldots, \underline{\mathbf{v}}_k)$ = subspace of vector space V spanned by vectors $\underline{\mathbf{v}}_i \in V$, ie. all linear combinations of the $\underline{\mathbf{v}}_i$, [10, p103]
- $\mathbb{F}^{m \times n}$ = $m \times n$ matrix over the field $\mathbb{F}$
- $\mathrm{CF}(A)$ = cofactor of matrix $A \in \mathbb{F}^{n \times n}$
- $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$ (2×2 determinant notation)
- $\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}$ (3×3 determinant notation)
- REF = row echelon form matrix, RREF = reduced row echelon form matrix
- ERO = elementary row operation on a matrix, EM = elementary matrix
- LI = linearly independent, LD = linearly dependent
- $\odot$ = pointing perpendicularly out of page, $\otimes$ = pointing perpendicularly into page
- $\binom{[n]}{k}$ = set of all subsets of $[n]$ of size k
- P_k = set of all bijections ('permutations') : $[k] \to [k]$, $P_{k,n}$ = set of all injections : $[k] \to [n]$, $(k \leq n)$
- Q_k = set of all mappings : $[k] \to [k]$, $Q_{k,n}$ = set of all mappings : $[k] \to [n]$
- $R_{k,n}$ = set of all strictly increasing injections : $[k] \to [n]$, $(k \leq n)$
- $(j_1, \ldots, j_k) \in Q_{k,n}$ = the mapping taking r to $j_r \in [n]$ for each $r \in [k]$
- $\mathrm{PERM}(n)$ = the set of all permutations of n objects
- S_n = the symmetric group of degree n consisting of all bijections from $[n]$ to $[n]$
- $\mathrm{sgn}(\sigma)$ = signature of permutation σ
- $\neg$ = logical 'not' symbol
- iff = if and only if

The properties of the 2×2 matrix inverse and determinant will be assumed in this article. The word 'category' used in this article refers to the common definition of the word rather than the formal mathematical notion.

2 Euclidean Space

> I am becoming increasingly convinced that the necessity of our geometry cannot be proven; at least not by a human mind or to the human mind ... geometry should not be classed with Arithmetic, which holds a priori, but with the likes of Mechanics. [1]
>
> - Carl Friedrich Gauss (1777-1855)

The following discussion gives a brief overview of some elementary notions of Euclidean geometry which although quite basic are worthwhile to review in a context which is independent of the notion of coordinate system. This is relevant to the geometric method of deriving the 3×3 matrix inverse described in section §7 since this method originates from consideration of how to transform coordinates between a standard Cartesian coordinate system and a general 'parallelopiped' or 'skew' coordinate system which is based on an arbitrary set of non-coplanar vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$. In order to derive such a coordinate transformation we need to initially view the geometric situation in a purely coordinate system free way. In §3 of this document we develop the properties of scalar product, vector product, scalar triple product, vector triple product, and the properties of right and left handedness in such a way, with coordinate systems being deferred until §4.

We do not discuss any formal set-theoretic foundations for Euclidean geometry but we do employ some basic concepts of 'naive' set theory [2] with a set simply being a well-defined collection of objects. Much of the commentary in this introduction is not required for matrix theory but the general geometric background is helpful in the more general sense to the process of mathematical thinking. The purely geometric aspect is often absent from contemporary sources which take a more analytical approach based on coordinate systems. In a sense the entire theory of the $n \times n$ matrix inverse derives from the two simple geometric concepts of projection and perpendicularity since these naturally lead to the solution of the 3×3 matrix inverse problem via their expressions in the scalar and vector product operations and this solution then readily generalizes to the $n \times n$ case.

The ideas below may be developed more fully and this would involve setting down some 'common notions' and 'axioms' and then formally making deductions from those to establish the basic properties needed. However for many purposes it suffices to just know in principle how that may be done and to have worked through proofs of some of the required theorems to get a general sense of how it works, without having to undertake the considerable amount of work that would be involved in a full formal development. This provides a sensible foundation and in the light of Gauss's quote above a realization that certain basic facts or 'common notions' need to be assumed or postulated just as in physics.

Euclidean geometry is in a sense the simplest branch of physics, being an exploration of the relationships and symmetries which arise within physical space, without introducing any other physical phenomena at all, and this then provides a foundation for mechanics where ideas like mass, time, motion, and force are introduced. As in any physics topic there is a firstly a real world phenomenon, then a physical model, and then a mathematical model which arises out of that. The mathematical model comprises formal mathematical objects such as numbers, sets, operators, functions, relations, and probabilities. Geometric ideas like point, line, plane, distance, direction, angle, area, volume, curve length and surface area belong to the physical model of Euclidean geometry and correspond directly with those concepts in the physical world. These concepts then link to the purely mathematical objects like number systems, angle measures, integrals and derivatives on $\mathbb{R}$ and $\mathbb{R}^n$, and abstract algebraic structures. We can introduce a set into a mathematical model wherever such a notion exists in the corresponding physical model and is a well defined collection of objects, so for example ES, EV, $\mathbb{A}$, and $\mathbb{D}$ defined below formally become sets.

At every stage in such a development we are either deducing something from earlier propositions or we are introducing an assumption which we have to make which will then be regarded as a 'postulate' or 'axiom', and there may be a number of different possible paths for such a process. Because of the myriad of 'common notions' that are required in geometry providing a discrete list of them all would not be a trivial task. And as in physics a measure of common sense is required in the study and application of this subject. An example postulate is the boundary property of $\mathbb{R}$ stated in Appendix C which is an intuitive assumption that we may reasonably make of the geometrically defined real number system — from this we are then able to deduce the important completeness property of $\mathbb{R}$, and indeed we can establish in an 'a priori' fashion the equivalence of the boundary property and the completeness property for any general totally ordered set $(X, \leq)$, [14].

For the purposes of this article we will consider Euclidean space (ES) to be the three-dimensional space of Euclidean geometry as studied in Euclid's Elements [3]. The term ES may refer to the overall concept of Euclidean space or it may refer to the set comprising all the points of Euclidean space. With the natural distance function d on ES, (ES, d) becomes a metric space ([4, p3-4], [6, p15]). We can denote the two and one dimensional versions of ES by ES_2 and ES_1 — these can be 'embedded' within the 3D ES.

Although it may seem strange we can have more than one ES within space itself. These correspond with different frames of reference in physics. For example we may have a fixed ground frame of reference and a second frame of reference attached to a rotating body, so that we have two ES's, one rotating with respect to the other. There does not need to be a coordinate system present for an ES to serve as a frame of reference, though clearly in practice we would often find it convenient to set up a coordinate system. The critical thing is that the ES expresses our basic geometric ideas which exist prior to the creation of coordinate systems. In the foregoing example, in the ES which is rotating a direction may be seen as fixed but in the ES of the ground frame that direction would be seen as changing. However if the second ES were moving purely translationally with respect to the first then the two ES's would have a common notion of direction throughout time.

The idea of distance is a numerical ratio of straight line segments with respect to some underlying unit of length that has been selected for the ES (with distance along curved line segments being defined using a limiting process of such straight line ratios). It requires the idea of a real number to express since as the length $\sqrt{2}$ of the diagonal of a unit square demonstrates distances in ES cannot always be expressed as rational numbers, ie. as ratios of two integers. The real number system $\mathbb{R}$ can in fact be *defined* in terms of this idea of numerical ratios of straight line segments in ES, by first constructing the real number line which is effectively a one dimensional ES. The first stage of that process is to designate a zero point and a 'positive' direction on the line and define the $\leq$ operator with respect to the positive direction to establish a total order relation [14] on $\mathbb{R}$, an order which is 'dense' in $\mathbb{R}$. The real numbers are then the points on that line. We can show by a geometric argument that this order relation $\leq$ has the 'completeness' property, ie. every non-empty subset which is bounded above has a least upper bound (Appendix C). We would then define a unit of distance on the line and with reference to that unit define the subsets $\mathbb{N}$, $\mathbb{W}$, $\mathbb{Z}$, $\mathbb{Q}$ within $\mathbb{R}$ and define the natural operations of addition, multiplication, substraction and division which make $(\mathbb{Q}, +, \cdot)$ a field that is 'dense' in $\mathbb{R}$. Using this density property of $\mathbb{Q}$ we can extend the definition of the field operations on $\mathbb{Q}$ to $\mathbb{R}$ to obtain $\mathbb{R}$ as an ordered field with $\mathbb{Q}$ as a subfield.

This establishes $(\mathbb{R}, +, \cdot\ , \leq)$ as a complete ordered field which is the basic algebraic structure studied in real analysis [8] and in which the concepts of calculus are formalized. A complete ordered field may be constructed by other means for example using Dedekind cuts or equivalence classes of Cauchy sequences in $\mathbb{Q}$, [15], however whichever method is used we end up with the same algebraic structure since it can be shown that any two complete ordered fields must be isomorphic. The real number line is sometimes referred to as the 'continuum', which reflects the fact that the set of points on this line are 'continuous' without gaps. This is because they have been defined in terms of the continuous line so there are no gaps by definition. The rational subset $\mathbb{Q}$ of this line however does have gaps for example the real number $\sqrt{2}$ which is the length of the diagonal of a unit square. $\sqrt{2}$ must belong to the line since the line contains all possible geometric ratios including those formed by limits of infinite sequences such as π, [44]. The completeness property of $\mathbb{R}$ is also known as the 'continuum property', the 'least upper bound property', the 'supremum property', and the 'Dedekind completeness property'. In Appendix C it is shown how to derive the completeness property of $\mathbb{R}$ using a geometric argument. When we come to study angles we regard these also as forming a continuum like the real numbers.

The idea of direction in a ES is a primitive concept that is not reducible to any simpler notion. It may be expressed in terms of a coordinate system, for example using Cartesian coordinates (x, y, z) or spherical coordinates (r, θ, φ), however the notion itself is not defined by coordinates. Given any point in ES we have the set $\mathbb{D}$ of all possible 3D directions emanating from that point. Although we cannot define what 'direction' actually is it is something that we are familiar with and we can make certain statements about it such as the following :

- two directions may coincide
- two directions may be opposite
- two directions may be parallel
- two non-parallel directions define a two dimensional plane
- three or more directions may be coplanar
- a certain angle exists between two directions

The concepts of point, distance and direction combine to produce the idea of a directed line segment (DLS) [13]. A DLS is defined as the straight line segment joining an ordered pair of two points in ES, the first being the starting point of the line segment and the second being the ending point of the line segment. The DLS thus has both a specific length and a specific direction. But two DLS's may have common length and direction yet still be distinct as they have different starting points. The DLS from point P to point Q is denoted $\overrightarrow{PQ}$.

The ideas of direction and DLS lead to the notion of a planar angle. Two non-zero non-parallel DLS's with common starting point define a plane and an angle between them. If we denote one of the DLS's as the first and the other as the second we can define angles with a sense of clockwise (CW) and angles with a sense of anticlockwise (ACW) as we view

the plane from a particular side. If we view the plane from the other side the angle sense flips between CW and ACW so angle sense is always relative to a particular side of the plane of the angle. If we change the order of the DLS's the angle sense also flips. In some cases we are only interested in the magnitude of the angle and not in the sense and in that case the order of the two DLS's that contain the angle is not important. We can allow for a zero angle which would not reside within any specific plane. Like direction the concept of angle is primitive and not reducible to any simpler notion and it is independent of angle measures like radians and degrees. As with direction though we can make statements about angles that we regard as true, ie. 'axioms' or 'postulates' of a kind :

- an angle may be zero, CW, or ACW
- an angle may be a 'right angle', or a 'straight angle'
- an angle may be a whole angle (ie. a full revolution)
- an angle may be an acute, obtuse or reflex angle
- an angle may be a 'wrap around' angle (ie. greater than a full revolution)
- one angle may be greater than or less than another angle
- an angle seen as CW from one side of the angle plane would be seen as as ACW from the other side

Although they are not numbers we can define a natural operation of addition on the set $\mathbb{A}$ of all angles by composing the turning effect of the angles. By convention ACW angles are regarded as positive and CW angles are regarded as negative. The additive inverse of an angle is the angle with the same magnitude but opposite sense. The magnitude $|A|$ of an angle A may be defined as itself if ACW or the additive inverse if CW. $(\mathbb{A}, +)$ then forms an abelian group with a natural order relation '$\leq$' which has the following properties in common with the order relation on $\mathbb{R}$:

- the order is a total order (ie. any two angles are comparable), [14]
- the order is compatible with the addition operation (ie. $A \leq B \Rightarrow A + C \leq B + C$)
- the order is dense in $\mathbb{A}$
- the magnitude operation satisfies the triangle inequality wrt to the order, ie. $|A + B| \leq |A| + |B|$
- the order satisfies the completeness property (Appendix C)

The completeness property of the order $\leq$ on $\mathbb{A}$ can be proved geometrically in a similar manner to $\mathbb{R}$ (Appendix C) since the angles form a continuum in the same way that $\mathbb{R}$ does, ie. there are by definition no gaps. We can then define sequence convergence in $\mathbb{A}$ exactly paralleling $\mathbb{R}$ ([8, p72-73]) : $\lim_{n \to \infty} A_n = A$ if $\forall\, \epsilon \in \mathbb{A}^+, \exists\, N \in \mathbb{N}$ such that $|A_n - A| < \epsilon\ \forall\, n > N$. By the triangle inequality such a limit is then unique when it exists. As in $\mathbb{R}$ ([8, p89] and [9, p111]) the Monotone Convergence Theorem can be deduced from the completeness property so that if (A_n) is an increasing sequence in $\mathbb{A}$ which is bounded above then it converges to $\sup\{A_n : n \in N\}$.

We would then define what we mean by multiplying or dividing an angle by an integer (positive or negative), and then by a rational number, and then using the density of the rationals within the reals, by a real number : eg. for $k \in \mathbb{R}^+$ and $A \in \mathbb{A}^+$, we could define $kA = \sup\{qA : 0 < q < k, q \in \mathbb{Q}\}$. For this purpose we would require to make some suitable postulates, for example :

- for any angle A and $n \in \mathbb{N}$ there is an angle exactly one n^{th} of A
- given any two angles A and B, with $B \neq 0$, there exists a real number k such that $A = kB$.

We would next want to define what we mean by an angle measure. A reasonable definition would be a function $m : \mathbb{A} \mapsto \mathbb{R}$ which maps $\mathbb{A}^+$ to $\mathbb{R}^+$ and which is additive, ie. $m(A+B) = m(A)+m(B)$. We could then deduce all the basic properties of m, such as $m(kA) = k\, m(A), \forall\ k \in \mathbb{R}$ and angles $A \in \mathbb{A}$. Additivity on the first two quadrants would suffice to prove m is an angle measure if we simply added integral multiples of a straight angle for angles outside of that range. We could postulate that given any angle $A \in \mathbb{A}^+$ and $x \in \mathbb{R}^+$ there is a unique angle measure m such that $m(A) = x$. Then if A is a straight angle degree measure would correspond with $x = 180$ and radian measure with $x = \pi$. A formal definition of radian measure is given in [44]. Using similar right-angled triangles the trigonometric functions $\sin, \cos, \tan$ are defined geometrically within ES as functions from $\mathbb{A}$ to $\mathbb{R}$ and they are linked via the radian angle measure to the analytically defined trigonometric functions [8, p315] which are functions from $\mathbb{R}$ to $\mathbb{R}$ defined using power series, so that if m is the radian measure and s is the analytic version of sin then we have $\mathrm{s}(m(A)) = \sin(A), \forall\ A \in \mathbb{A}$.

It is important to understand the native physical model notions in Euclidean geometry prior to their representations with numbers such as coordinate systems, angle measures, and real numbers. This is because the numerical representations can change but the fundamental ideas do not. For example whether we measure them in degrees or radians

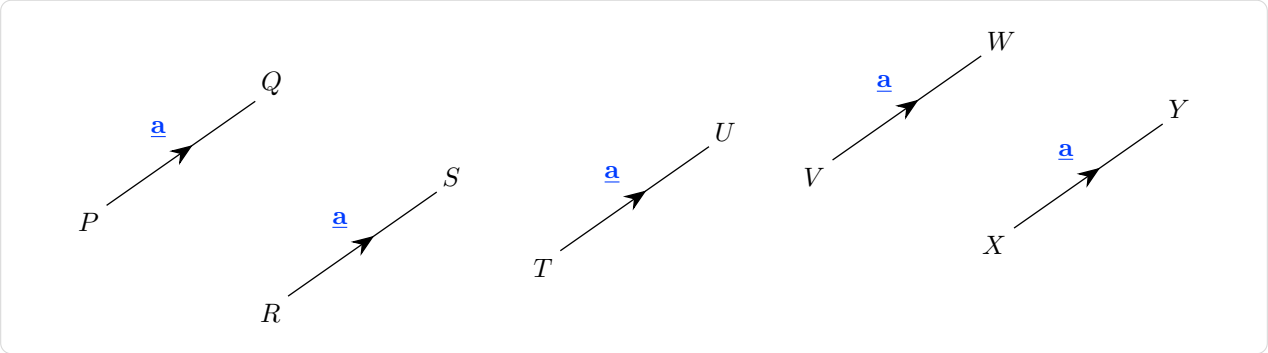
angles are still the same fundamental idea possessing the same relationships and properties and likewise direction is a concept independent of the coordinate system used to express it. On the real number line we were able to establish the completeness property of the total order relation even before introducing the concept of a natural number onto the line (Appendix C), ie. whilst the 'real number' was still a point on the line rather than a 'number' as we would normally understand it, and this proof method carried over to the completeness property for the set of angles $\mathbb{A}$. In the case of area, volume, curve length, and surface area we must have a native idea of what these concepts mean geometrically since when we express them formally using mathematical analysis different expressions result depending on the coordinate system we select, for example we could rotate a Cartesian coordinate system in 3D in an arbitrary manner [45] or we could replace it with a spherical coordinate system [46], [47]. Theorems in mathematical analysis (eg. Transformation Of Integrals Theorem [9, §24.26, p335]) prove that the analytic definitions of these quantities are consistent and independent of the coordinate system we choose.

In the present problem, asking what the consequence is of changing coordinate systems leads to a simple geometric solution to the 3×3 case of the real matrix inverse (§7) which then naturally generalizes to the $n \times n$ case over a general field (§12) using permutations [43]. So we are starting out with geometric reasoning, which as in the Gauss quote above is a form of physics, but then we can use the insights gained from that in a purely 'a priori' formal system of algebraic reasoning that in itself does not depend on any geometric intuitions. In Euclid's Elements [3] representations by coordinate systems, angle measures, real numbers and algebra are not used, only the native geometric notions are used. We can still have the benefit of the geometric insight of Euclid without sacrificing the rigorous proof of modern mathematics.

3 Euclidean Vectors

It is convenient to group together all DLS's which have common length and direction and this leads to the idea of a *vector*. An equivalence relation called equipollence [16] is defined on the set of all DLS's in ES so that two DLS's are deemed equipollent if they have common length and direction. We can then partition the set of all DLS's into the equivalence classes of this relation, so like identical arrows being fired in parallel all DLS's with common length and direction would be grouped together irrespectively of their starting points (Figure 1). Thus each equivalence class would be associated with a specific length and direction. We call this combination a vector, ie. an ordered pair (l, d) where l is a non-negative real number representing the length and $d \in \mathbb{D}$, the set of directions. There is a special type of vector called the zero vector which is of zero length and its direction is 'undefined', and it is represented not with an ordered pair but just with the number zero. The length of a vector is also sometimes called a 'magnitude', since vectors are used in physics to represent a wide variety of physical quantities such as force, velocity, electric field, and so on.

Figure 1: Equipollent DLS's in ES each corresponding to the same vector $\underline{\mathbf{a}}$

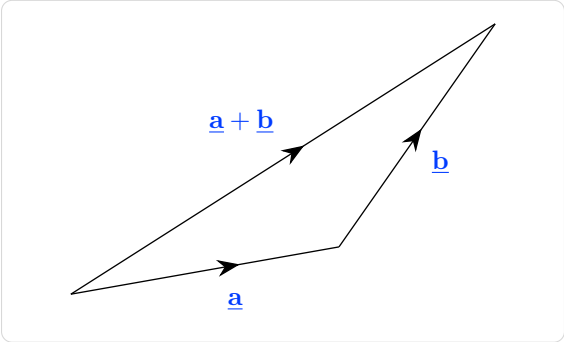


We define the set Euclidean Vectors (EV) to consist of all possible vectors in ES. Every distinct ordered pair (l, d) corresponds to a unique vector in EV, with additionally the zero vector which is not an ordered pair but just a length of zero with no direction defined. It is important to note that unlike a DLS a vector does not have a starting or ending point, it does not have any particular position within the ES. It can thus be 'placed' at any location within the ES, which effectively means we consider the DLS for the vector which is at that particular location. We nonetheless still just use the term 'vector' whether it has been 'placed' at a certain location (ie. it has been 'bound') or it remains a 'free vector' not associated with any specific location, the context will normally make it clear which of these two meanings is intended. We denote the two and one dimensional versions of EV by EV_2 and EV_1 — these can be embedded within EV.

A natural operation of addition can be made on EV by placing the vectors 'head to tail' (Figure 2), and a natural operation of scalar multiplication by a real number can be defined, so that EV forms a real vector space [10, p95].

EV has a norm which is the vector length $|\underline{\mathbf{a}}|$, wrt which the vector addition operation obeys the 'triangle inequality': $|\underline{\mathbf{a}} + \underline{\mathbf{b}}| \leq |\underline{\mathbf{a}}| + |\underline{\mathbf{b}}|$.

Figure 2: Vector addition in EV



Two non-parallel vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$ placed tail to tail at a point P in ES define a plane Γ in ES passing through P. There are two unsigned planar angles in Γ between $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$, called the minor angle in the range $(0°, 180°)$ and the major angle in the range $(180°, 360°)$. These angles can be made signed (ie. ACW or CW) if we specify which side of the plane we are viewing from and whether the angle is going from $\underline{\mathbf{a}}$ to $\underline{\mathbf{b}}$ or from $\underline{\mathbf{b}}$ to $\underline{\mathbf{a}}$. Two parallel vectors do not define a plane and have a single unsigned angle between them which is either $0°$ or $180°$.

Although a vector in EV comprises the properties of length and direction the general concept of a vector space V over a field F (eg. see [10, p95]) defines neither length nor direction of a vector. The purpose of a general vector space is to study the algebraic properties of the operations of addition and scalar multiplication on vectors. Vector spaces may be equipped with a norm [17] defining a length function on the vectors (as EV) but this is not defined within a general vector space.

3.1 Relationship Between EV And ES

Vectors in EV can be placed in one-to-one correspondence with points in ES by selecting an origin point O in ES and mapping a point $P \in ES$ to its 'position vector' wrt O, ie. the unique vector corresponding to the directed line segment $\overrightarrow{OP}$. Any non-zero vector defines the set of lines in ES which are parallel to it and exactly one of these lines passes through O. Given any two non-zero non-parallel vectors, placing their tails at O defines a unique plane in ES through O and then all the planes in ES parallel to that plane also 'contain' these two vectors (recall we can 'move' a vector to any position within ES but its length and direction must always remain the same). Any single non-zero vector also defines a unique plane in ES through O by acting as its normal, and then that vector will also be normal to all planes parallel to that plane. In short we are free to shift planes, lines, and DLS's anywhere within ES purely translationally (but not rotationally in any manner) and their relationships with 'free' (ie. 'unbound') vectors will always remain the same. Operations on DLS's correspond with operations on vectors, eg. in the case of addition, $\overrightarrow{PQ} = \overrightarrow{PR} + \overrightarrow{RQ}$ implies a vector relationship of form $\underline{\mathbf{a}} + \underline{\mathbf{b}} = \underline{\mathbf{c}}$ and vice-versa.

With these understandings we can freely interchange the terms vector, DLS, and point, and associate sets of parallel lines with single non-zero vectors and sets of parallel planes with pairs of non-zero non-parallel vectors, or single non-zero vectors as their normals, the specific meaning intended being determined by the context. For example we could write something like $\underline{\mathbf{a}} = \overrightarrow{PQ}$ to assign vector $\underline{\mathbf{a}}$ to be the vector associated with the directed line segment $\overrightarrow{PQ}$. A set of vectors of form $\Gamma = \{k\,\underline{\mathbf{a}} + l\,\underline{\mathbf{b}} : k, l \in \mathbb{R}\}$ for non-zero non-parallel vectors $\underline{\mathbf{a}}, \underline{\mathbf{b}} \in$ EV could be referred to as a plane, or a 'vector plane', since it corresponds with the unique plane through the origin in ES containing the vectors $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ and also with all planes in ES parallel to that plane since these two vectors 'lie within' all such planes. In the language of Linear Algebra the set Γ is a two-dimensional subspace of the vector space EV. Similarly $L = \{k\,\underline{\mathbf{a}} : k \in \mathbb{R}\}$ for non-zero vector $\underline{\mathbf{a}}$ is a one-dimensional subspace of EV and could be referred to as a line, or a 'vector line', since it corresponds with a unique line through the origin in ES defined by vector $\underline{\mathbf{a}}$ and also with all lines in ES parallel to that line.

A specific plane Γ in ES can be described by the set of position vectors of points on Γ wrt some origin O selected within the ES : ie. we can write $\Gamma = \{\underline{\mathbf{r}} \in \text{EV} : (\underline{\mathbf{r}} - \underline{\mathbf{r}}_0) \cdot \underline{\mathbf{n}} = 0\}$, where $\underline{\mathbf{n}}$ is a normal vector for the plane and $\underline{\mathbf{r}}_0$ is the position vector wrt O of some point within the plane (see §3.2 below for the scalar product operation). Likewise a specific line in ES can be described in terms of position vectors wrt O by a set in EV of form $\{\underline{\mathbf{r}}_0 + k\underline{\mathbf{a}} : k \in \mathbb{R}\}$ where $\underline{\mathbf{r}}_0$ is a point on the line and $\underline{\mathbf{a}}$ is a vector parallel to the line.

We use the term 'coplanar' to mean a given set of three or more vectors all lie within a common plane and include within this the case where the vectors are collinear (ie. parallel).

Note the distinction algebraically that EV is a vector space whilst ES is a metric space. There is no addition nor scalar multiplication operation defined on ES, however ES is the structure within which geometric concepts are visualized. An object of EV (ie. a 'free vector') can be placed into ES at an arbitrary location, where it then becomes a 'bound vector', ie. a DLS.

3.2 Scalar Product In EV

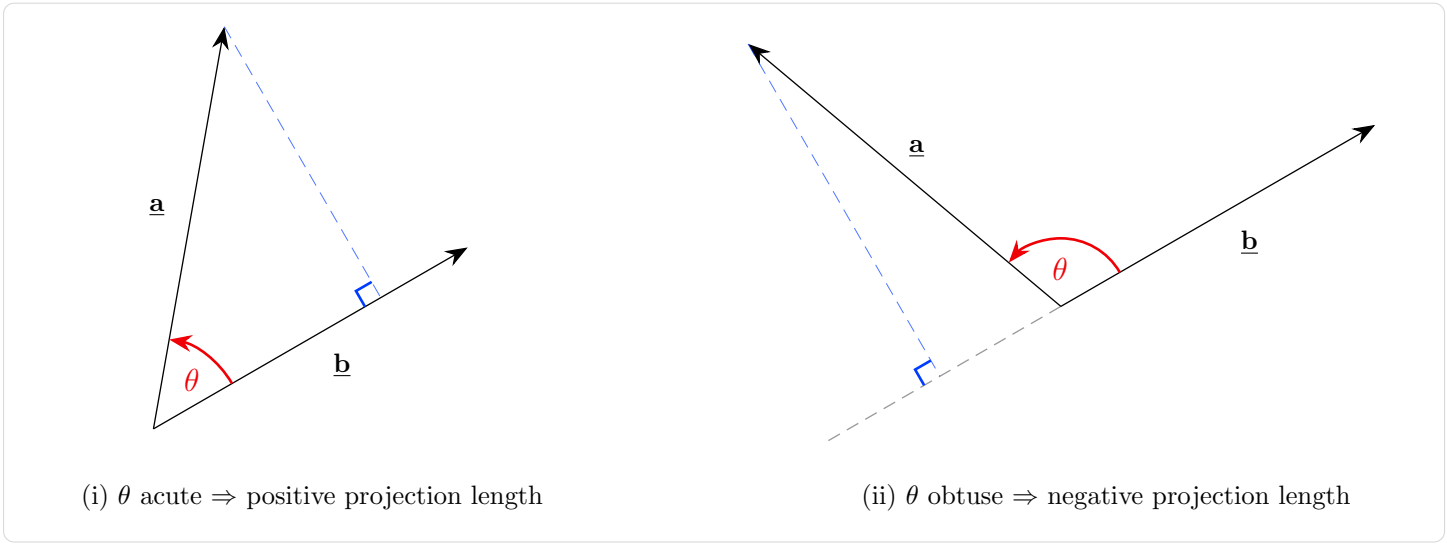
The scalar product (or dot product) $\underline{\mathbf{a}} \cdot \underline{\mathbf{b}}$ of vectors $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ in EV is defined to be the real number given by :

$$\underline{\mathbf{a}} \cdot \underline{\mathbf{b}} = |\underline{\mathbf{a}}|\,|\underline{\mathbf{b}}| \cos\theta$$

where θ is the angle between $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ when these are placed 'tail to tail'. Since $\cos(360° - \theta) = \cos\theta$ it doesn't matter whether θ is the minor angle or the major angle between $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$, but often we use the minor angle, though in some cases it is convenient to use the major angle. Also it doesn't matter which direction, ACW or CW, we take the angle in, ie. whether the angle goes from $\underline{\mathbf{a}}$ to $\underline{\mathbf{b}}$ or from $\underline{\mathbf{b}}$ to $\underline{\mathbf{a}}$, since $\cos(-\theta) = \cos\theta$, ie. the angle can be unsigned. Placing the vectors 'head to tail' though as in the definition of vector addition would give an angle of $180° - \theta$ and hence a different result, since $\cos(180° - \theta) = -\cos\theta$, so they must always be placed tail to tail for scalar product as in Figure 3.

From Figure 3 we see that if $\underline{\mathbf{b}}$ is a unit vector then $\underline{\mathbf{a}} \cdot \underline{\mathbf{b}}$ equals the length $|\underline{\mathbf{a}}| \cos\theta$ of the projection of $\underline{\mathbf{a}}$ onto $\underline{\mathbf{b}}$ (ie. the length of the component of $\underline{\mathbf{a}}$ that is in the direction of $\underline{\mathbf{b}}$). This is a signed length since it becomes negative when θ is obtuse, so that the component of $\underline{\mathbf{a}}$ in the direction of $\underline{\mathbf{b}}$ is the vector $(\underline{\mathbf{a}} \cdot \underline{\mathbf{b}})\,\underline{\mathbf{b}}$. Generally if $\underline{\mathbf{b}}$ is not necessarily a unit vector the component of $\underline{\mathbf{a}}$ in the direction of $\underline{\mathbf{b}}$ is $\dfrac{1}{|\underline{\mathbf{b}}|^2}\,(\underline{\mathbf{a}} \cdot \underline{\mathbf{b}})\,\underline{\mathbf{b}}$ and the scalar product of $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ is the signed projection length of $\underline{\mathbf{a}}$ onto $\underline{\mathbf{b}}$ times the length of $\underline{\mathbf{b}}$.

Figure 3: Projection of $\underline{\mathbf{a}}$ onto $\underline{\mathbf{b}}$



(i) θ acute $\Rightarrow$ positive projection length

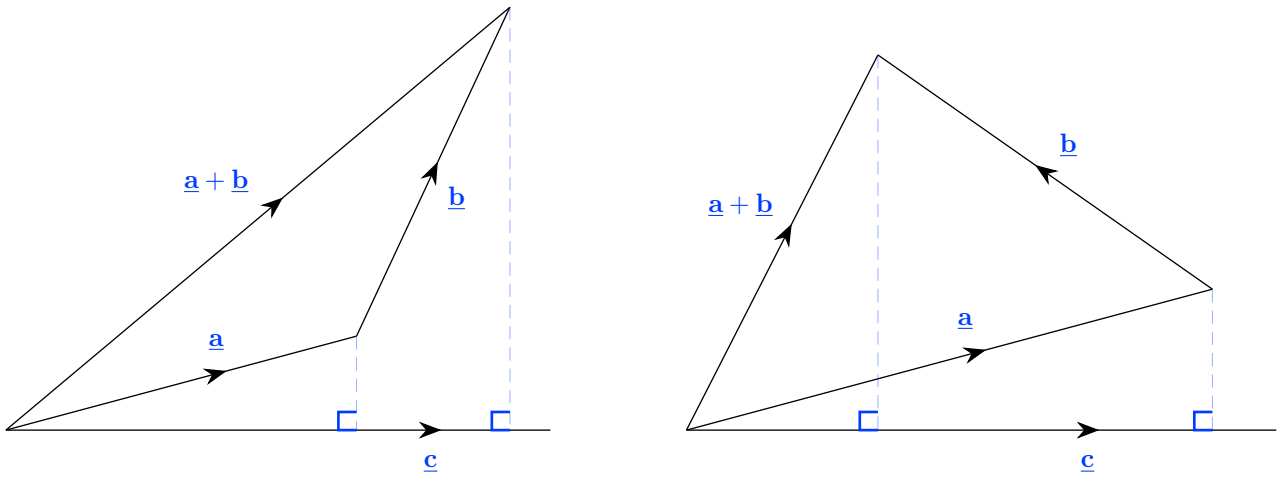
(ii) θ obtuse $\Rightarrow$ negative projection length

Scalar product has the following properties :

1. $\underline{\mathbf{a}} \cdot \underline{\mathbf{b}} = \underline{\mathbf{b}} \cdot \underline{\mathbf{a}}$ (commutative)
2. $\lambda(\underline{\mathbf{a}} \cdot \underline{\mathbf{b}}) = (\lambda\underline{\mathbf{a}}) \cdot \underline{\mathbf{b}} = \underline{\mathbf{a}} \cdot (\lambda\underline{\mathbf{b}})$ (multiplicative)
3. for non-zero $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$, $\underline{\mathbf{a}} \cdot \underline{\mathbf{b}} = 0 \iff \underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ are perpendicular
4. $(\underline{\mathbf{a}} + \underline{\mathbf{b}}) \cdot \underline{\mathbf{c}} = \underline{\mathbf{a}} \cdot \underline{\mathbf{c}} + \underline{\mathbf{b}} \cdot \underline{\mathbf{c}}$ (distributive)

The first three of these are clear from the definition. The distributive law holds since the projection of a sum is just sum of the projections, as shown in Figure 4.

Figure 4: Scalar Product Distributive Law



(i) two positive projections

(ii) a positive and a negative projection

3.2.1 EV As Direct Sum Of Line And Plane

In Linear Algebra the term 'direct sum' means to express a vector space as a 'sum' of subspaces [10, §43, p126], so each vector is a unique sum of vectors drawn from the respective subspaces. In the case of two subspaces this requires the subspaces to have intersection $\{\underline{\mathbf{0}}\}$. The vector space EV can be expressed as a direct sum of any line and plane provided the line is not parallel to the plane. We can derive an expression for the unique coefficients of a general vector $\underline{\mathbf{r}} \in$ EV wrt such a line and plane in terms of the scalar product operation.

Theorem 1 (Direct Sum Of Line And Plane). Given a plane Γ in ES with a normal vector $\underline{\mathbf{n}}$, and a line L in ES parallel to a vector $\underline{\mathbf{a}}$, then if L is not parallel to Γ then any vector $\underline{\mathbf{r}} \in$ EV can be expressed in a unique way in the form :

$$\underline{\mathbf{r}} \;=\; \lambda\,\underline{\mathbf{a}} + \underline{\mathbf{b}}$$

where λ is a real number and $\underline{\mathbf{b}}$ a vector parallel to Γ, the unique values of λ and $\underline{\mathbf{b}}$ being given by :

$$\lambda \;=\; \frac{(\underline{\mathbf{r}} \cdot \underline{\mathbf{n}})}{(\underline{\mathbf{a}} \cdot \underline{\mathbf{n}})} \tag{1}$$

$$\underline{\mathbf{b}} \;=\; \underline{\mathbf{r}} - \left(\frac{\underline{\mathbf{r}} \cdot \underline{\mathbf{n}}}{\underline{\mathbf{a}} \cdot \underline{\mathbf{n}}}\right) \underline{\mathbf{a}} \tag{2}$$

Proof. Let P be the point of intersection of L and Γ and place $\underline{\mathbf{r}}$ with its tail at point P as shown in Figure 5. Let Q be the head of $\underline{\mathbf{r}}$ and let Δ be the plane parallel to Γ which contains Q. Let R be the point where L intersects plane Δ. Then :

$$\begin{aligned} \underline{\mathbf{r}} &= \overrightarrow{PQ} \\ &= \overrightarrow{PR} + \overrightarrow{RQ} \\ &= \lambda\,\underline{\mathbf{a}} + \underline{\mathbf{b}} \end{aligned}$$

for some $\lambda \in \mathbb{R}$ and vector $\underline{\mathbf{b}}$ parallel to plane Γ, since $\overrightarrow{PR}$ lies on line L which is parallel to vector $\underline{\mathbf{a}}$, and $\overrightarrow{RQ}$ lies in plane Δ which is parallel to plane Γ. Note how we can freely interchange vectors and DLS's, as described in §3.1.

If we had some other expression of $\underline{\mathbf{r}}$ in this form, say $\underline{\mathbf{r}} = \lambda'\,\underline{\mathbf{a}} + \underline{\mathbf{b}}'$, then if $\lambda \neq \lambda'$:

$$\begin{aligned} \lambda\,\underline{\mathbf{a}} + \underline{\mathbf{b}} &= \lambda'\,\underline{\mathbf{a}} + \underline{\mathbf{b}}' \\ \Rightarrow\; \underline{\mathbf{a}} &= \left(\frac{1}{\lambda - \lambda'}\right)(\underline{\mathbf{b}}' - \underline{\mathbf{b}}) \end{aligned}$$

$\Rightarrow$ $\underline{\mathbf{a}}$ is parallel to plane Γ, since $\underline{\mathbf{b}}'$ and $\underline{\mathbf{b}}$ are parallel to Γ

$\Rightarrow$ contradiction since line L is not parallel to Γ.

Thus $\lambda' = \lambda$ and then $\underline{\mathbf{b}}' = \underline{\mathbf{b}}$, establishing the uniqueness of λ and $\underline{\mathbf{b}}$. To obtain the value of λ — since $\underline{\mathbf{b}}$ is parallel to Γ and $\underline{\mathbf{n}}$ is a normal to Γ we have :

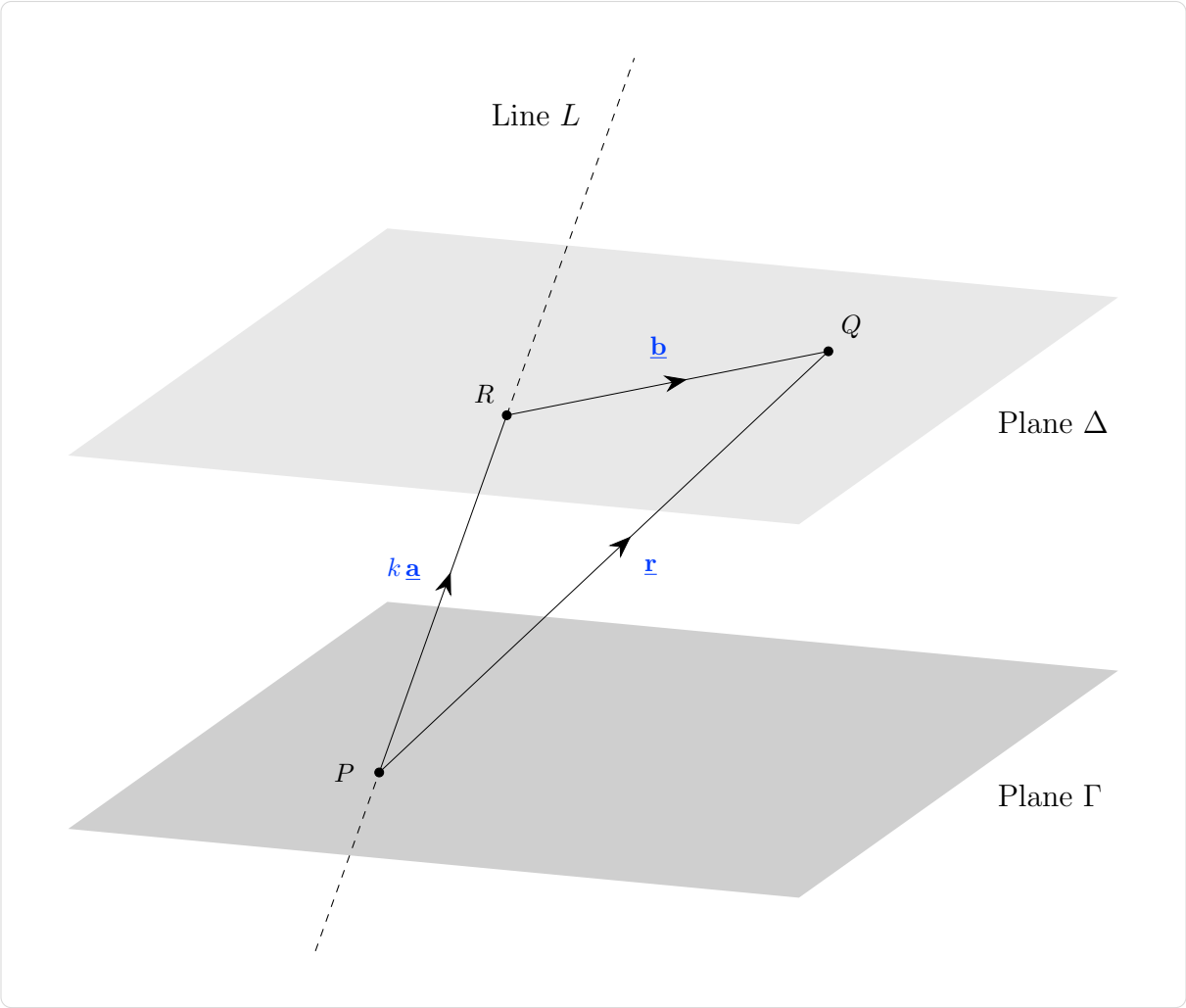
$$\begin{aligned} \underline{\mathbf{b}} \cdot \underline{\mathbf{n}} &= 0 \\ \text{so} \quad (\underline{\mathbf{r}} - \lambda\,\underline{\mathbf{a}}) \cdot \underline{\mathbf{n}} &= 0 \\ \text{thus} \quad \underline{\mathbf{r}} \cdot \underline{\mathbf{n}} &= \lambda\,\underline{\mathbf{a}} \cdot \underline{\mathbf{n}} \end{aligned}$$

from which the required expressions for λ and then $\underline{\mathbf{b}}$ follow.

Note that neither $\underline{\mathbf{n}}$ nor $\underline{\mathbf{a}}$ need to be unit vectors in the above proof.

QED

Figure 5: EV As Direct Sum Of Line L And Plane Γ



3.2.2 EV_2 As Direct Sum Of Two Lines

Similarly to §3.2.1 above we can use the scalar product to express a vector plane EV_2 as a direct sum of two non-parallel lines. Note the normals $\underline{\mathbf{m}}$, $\underline{\mathbf{n}}$ in Theorem 2 below must not go outwith the plane EV_2, otherwise the theorem will not be valid.

Theorem 2 (Direct Sum Of Two Lines). Given non-parallel lines L_1 and L_2 in ES_2, which are parallel respectively to vectors $\underline{\mathbf{a}}, \underline{\mathbf{b}} \in \mathrm{EV}_2$, then if $\underline{\mathbf{m}}, \underline{\mathbf{n}} \in \mathrm{EV}_2$ are normals to $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$ respectively, then any vector $\underline{\mathbf{r}} \in \mathrm{EV}_2$ can be expressed in the form :

$$\underline{\mathbf{r}} \quad = \quad \lambda\,\underline{\mathbf{a}} + \mu\,\underline{\mathbf{b}}$$

where the unique values of $\lambda, \mu \in \mathbb{R}$ are given by :

$$\lambda = \frac{(\underline{\mathbf{r}} \cdot \underline{\mathbf{n}})}{(\underline{\mathbf{a}} \cdot \underline{\mathbf{n}})} \tag{3}$$

$$\mu = \frac{(\underline{\mathbf{r}} \cdot \underline{\mathbf{m}})}{(\underline{\mathbf{b}} \cdot \underline{\mathbf{m}})} \tag{4}$$

Proof. Let P be the point of intersection of L_1 and L_2, and place $\underline{\mathbf{r}}$, $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$ tail to tail at point P and let Q be the head of $\underline{\mathbf{r}}$ as shown in Figure 6. Let the line through Q parallel to L_2 meet L_1 at point R. Then

$$\begin{aligned} \underline{\mathbf{r}} &= \overrightarrow{PQ} \\ &= \overrightarrow{PR} + \overrightarrow{RQ} \\ &= \lambda\,\underline{\mathbf{a}} + \mu\,\underline{\mathbf{b}}, \end{aligned}$$

for some $\lambda, \mu \in \mathbb{R}$ since $\overrightarrow{PR}$ lies in line L_1 and $\overrightarrow{RQ}$ lies in a line parallel to L_2.

If we had another expression of this form, say :

$$\underline{\mathbf{r}} = \lambda'\,\underline{\mathbf{a}} + \mu'\,\underline{\mathbf{b}}\,,$$

with $\lambda \neq \lambda'$ then :

$$\begin{aligned} \lambda\,\underline{\mathbf{a}} + \mu\,\underline{\mathbf{b}} &= \lambda'\,\underline{\mathbf{a}} + \mu'\,\underline{\mathbf{b}} \\ \Rightarrow\ \underline{\mathbf{a}} &= \left(\frac{\mu' - \mu}{\lambda - \lambda'}\right)\underline{\mathbf{b}} \end{aligned}$$

$\Rightarrow$ contradiction since $\underline{\mathbf{a}}$ is not parallel to $\underline{\mathbf{b}}$.

Thus $\lambda = \lambda'$ and then $\mu = \mu'$, establishing uniqueness of λ, μ. To obtain the value of λ :

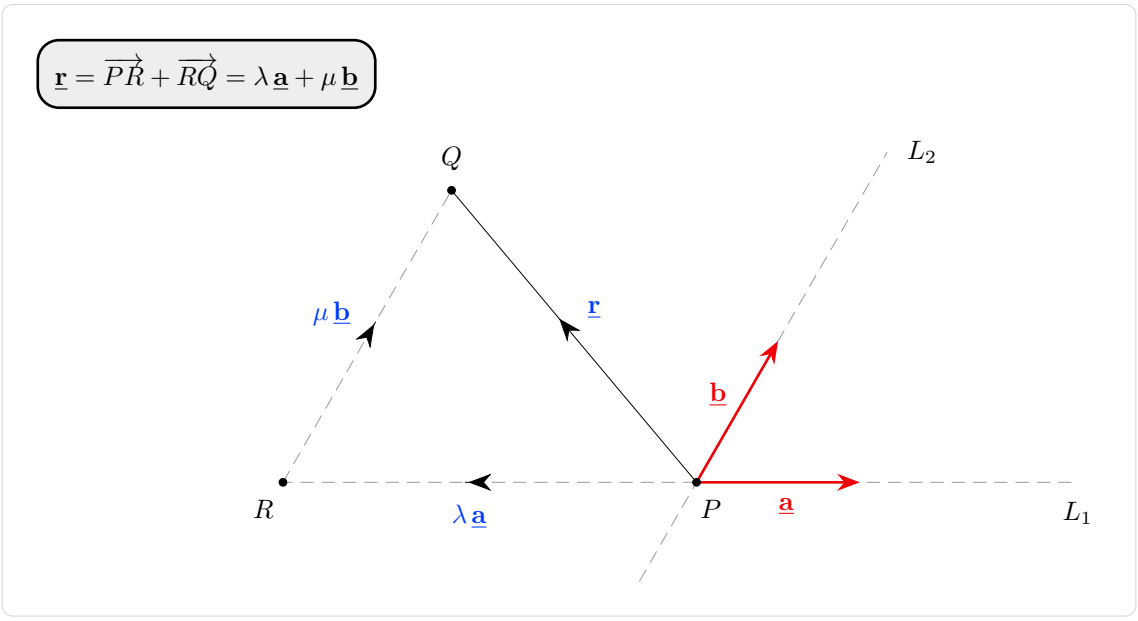
$$\begin{aligned} \underline{\mathbf{r}} \cdot \underline{\mathbf{n}} &= \lambda\,\underline{\mathbf{a}} \cdot \underline{\mathbf{n}}\,, \qquad \text{since } \underline{\mathbf{n}} \text{ is normal to } \underline{\mathbf{b}} \\ \Rightarrow\ \lambda &= \frac{(\underline{\mathbf{r}} \cdot \underline{\mathbf{n}})}{(\underline{\mathbf{a}} \cdot \underline{\mathbf{n}})} \end{aligned}$$

since $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{n}}$ all lie in a common plane, $\underline{\mathbf{n}}$ is normal to $\underline{\mathbf{b}}$, and $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$ are non-parallel, so that $\underline{\mathbf{n}}$ cannot also be normal to $\underline{\mathbf{a}}$, ie. $\underline{\mathbf{a}} \cdot \underline{\mathbf{n}} \neq 0$. The value for μ is obtained by similar argument using the normal $\underline{\mathbf{m}}$ to $\underline{\mathbf{a}}$.

Note the normal vectors $\underline{\mathbf{m}}$, $\underline{\mathbf{n}}$ do not need to be unit vectors in the above proof.

QED

Figure 6: EV_2 As Direct Sum Of Two Lines L_1 And L_2



3.3 Right-Hand Rule In EV

In §2 we discussed how an ordered pair of non-zero non-parallel vectors, identified with DLS's, and placed tail to tail defines a signed angle within the plane contained by the vectors. It is convenient to define a notation for this plane.

Definition 1. $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ is defined to be the plane containing the non-zero non-parallel vectors $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$.

□

$\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ is actually a collection of infinitely many parallel planes in ES, but for the purposes of defining angles it doesn't matter which one of these we select. The ACW angles were designated as 'positive' and the CW angles as 'negative' — however we noted that this sign was flipped if we viewed from the other side of the plane $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ or if we changed the order of the vectors $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$.

The idea of the 'right hand rule' is to designate a particular side $\mathrm{RH}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ of the plane $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ as being the 'right-handed side'. (By 'particular side' we mean one half of the ES divided by the plane). There are two (unsigned) angles between $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$, the minor angle in the range $(0°, 180°)$ and the major angle in the range $(180°, 360°)$ and these become signed angles if we specify the direction of the angle as being from $\underline{\mathbf{a}}$ to $\underline{\mathbf{b}}$ — the sense of this direction being dependent upon which side of the plane $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ we view from. In particular the sign of the minor angle from $\underline{\mathbf{a}}$ to $\underline{\mathbf{b}}$ must be ACW when viewed from one side of the plane $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ and CW when viewed from the other side of $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$.

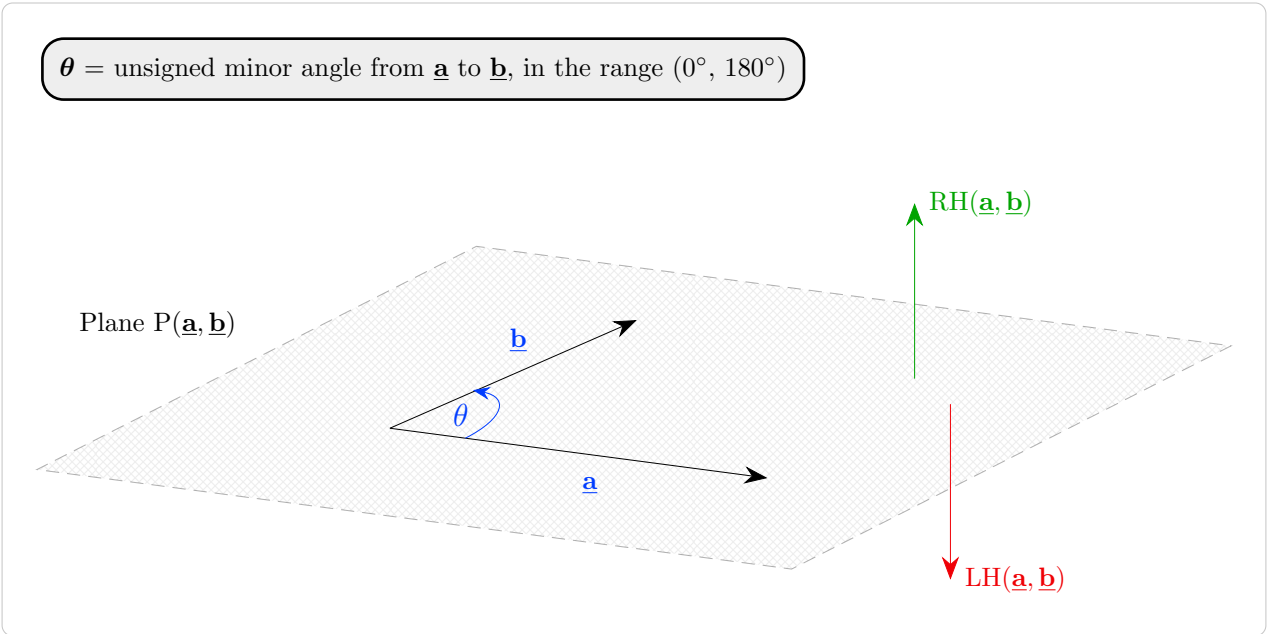
Definition 2. $\mathrm{RH}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ is defined to be the side of the plane $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ which produces an ACW sense for the minor angle from $\underline{\mathbf{a}}$ to $\underline{\mathbf{b}}$ when viewed from that side. $\mathrm{LH}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ is defined to be the opposite side of $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ from $\mathrm{RH}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$.

□

As seen in Figure 7, $\mathrm{RH}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ is the same side of $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ pointed to by our thumb if we curl the fingers of our right hand around the minor angle from $\underline{\mathbf{a}}$ to $\underline{\mathbf{b}}$, hence the term 'right-hand rule'. $\mathrm{LH}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ is similarly determined by the same procedure with our left hand.

A 'right hand rule' convention also exists for the sign of an angle of rotation about a directed axis in EV. This assigns a positive sign to a rotation angle if when we view the rotation with the axis pointing towards us it is ACW, and a negative sign if it is CW. This convention is used in rotating a vector about a directed axis using the Rodrigues rotation formula [48, Appendix D].

Figure 7: Right And Left Hand Rules



With the above definition it is clear that $\mathrm{RH}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ is the opposite side from $\mathrm{RH}(\underline{\mathbf{b}}, \underline{\mathbf{a}})$, so that the order of the vectors $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ matters. However the plane $\mathrm{P}(\underline{\mathbf{b}}, \underline{\mathbf{a}})$ is the same plane as $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$. Neither $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ nor $\mathrm{RH}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ are defined if $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ are parallel or if one of them is zero.

Definition 3. An ordered list of three non-zero non-coplanar vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, and $\underline{\mathbf{c}}$ form a right-handed system if $\underline{\mathbf{c}}$ points to the right-handed side $\mathrm{RH}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ of plane $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$, otherwise they form a left-handed system.

□

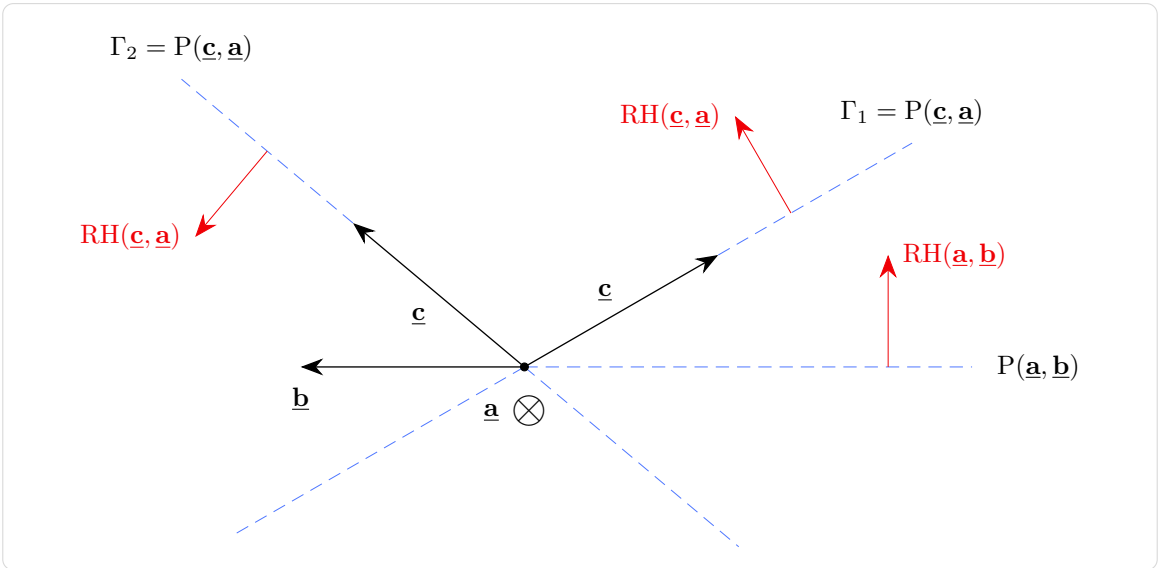
Note Definition 3 is dependent upon the order of the vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$. However we are able to show below in Theorem 3 that these properties are invariant under any cyclic reordering of $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$. Cycling three places forwards or three places backwards produces the original order and cycling two places forward is equivalent to cycling one place back, and cycling two places backward is equivalent to cycling one place forward. A cyclic reordering could never be the equivalent of a swap of two of the three objects since that would imply the cyclic reordering was non-identity but produced the identity when applied twice in succession, which is impossible since a forward cycle of one place applied twice gives a forward cycle of two places, and a forward cycle of two places applied twice gives a forward cycle of one place.

Theorem 3 (Preservation Of Handedness Under Cycling). If non-zero non-coplanar vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ form a RH system then these remain a right-handed system after cycling one place forward. Hence $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ remains a RH system after any cyclic reordering of $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$. Likewise a LH system remains a LH system after any cyclic reordering of the vectors. Thus the property of 'handedness' is preserved under cycling.

Proof. Consider a RH system $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ as shown in Figure 8, where $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ is perpendicular to the page, $\underline{\mathbf{a}}$ is directed into the page, the projection of $\underline{\mathbf{b}}$ onto the page is seen to the left of $\underline{\mathbf{a}}$, and the minor angle from $\underline{\mathbf{a}}$ to $\underline{\mathbf{b}}$ is to the left of $\underline{\mathbf{a}}$. Since $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ is right-handed, the projection of $\underline{\mathbf{c}}$ onto the page must be on the right-handed side $\mathrm{RH}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ of plane $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$, ie. pointing to the upper side of plane $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ in the diagram. Two possible positions of $\underline{\mathbf{c}}$ are shown forming planes Γ_1 and Γ_2 with vector $\underline{\mathbf{a}}$. Note $\underline{\mathbf{b}}$ and $\underline{\mathbf{c}}$ are not necessarily in the plane of the page but are shown as their projections into the page, so they may be pointing into or out of the page. From the diagram, considering all possible positions of planes such as Γ_1 and Γ_2, it is clear that however $\underline{\mathbf{c}}$ is situated above $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$, $\underline{\mathbf{b}}$ must always point to the side $\mathrm{RH}(\underline{\mathbf{c}}, \underline{\mathbf{a}})$ of $\mathrm{P}(\underline{\mathbf{c}}, \underline{\mathbf{a}})$ (though note this would not be the case if $\underline{\mathbf{c}}$ pointed below $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$). Thus according to the above definition $\underline{\mathbf{c}}$, $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$ must be a RH system. Thus cycling $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ one step forward preserves the right handedness of the system. Then any cycling of $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ preserves the right handedness as it is just a repetition of 1-cycles. If a cycle of a LH system resulted in a RH system then by cycling that RH system further we would arrive back at the LH system which is impossible due to the foregoing preservation of right handedness. Hence left handedness is also preserved under cycling.

QED

Figure 8: Cycling RH System Of Vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$



From the definition above it is clear that if we interchange the order of $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ then the vector triple $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ changes its 'handedness' from right to left or vice-versa, since $\mathrm{RH}(\underline{\mathbf{b}}, \underline{\mathbf{a}})$ is always on the opposite side of $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ from $\mathrm{RH}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$. If we interchanged $\underline{\mathbf{b}}$ and $\underline{\mathbf{c}}$, since this would be equivalent to cycling back one place, swapping the first two, and then cycling forward one place, the handedness would again be flipped, by Theorem 3. And if we interchanged $\underline{\mathbf{a}}$ and $\underline{\mathbf{c}}$ since this would be equivalent to cycling one place forwards, swapping the first two, then cycling one place backwards, the handedness would once more be flipped. So swapping any two of the vectors flips the handedness. The three possible orderings produced by these swaps are all cyclically related, ie. they can all be obtained from one another by a cyclic reordering.

Thus in summary the six possible reorderings of $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ divide into two groups of three, with each group being a '3-cycle' containing three cyclically related orderings all with the same handedness, and the two groups having opposite handedness :

1. the cyclic reorderings of the original order :
   (i) $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$
   (ii) $\underline{\mathbf{c}}$, $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$
   (iii) $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$, $\underline{\mathbf{a}}$
2. the reorderings obtained from swaps applied to the original order :
   (iv) $\underline{\mathbf{b}}$, $\underline{\mathbf{a}}$, $\underline{\mathbf{c}}$ (swap the first two)
   (v) $\underline{\mathbf{a}}$, $\underline{\mathbf{c}}$, $\underline{\mathbf{b}}$ (swap the second two)
   (vi) $\underline{\mathbf{c}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{a}}$ (swap the first and third)

Performing a swap on any of these six reorderings will send it to the other group and flip the handedness.

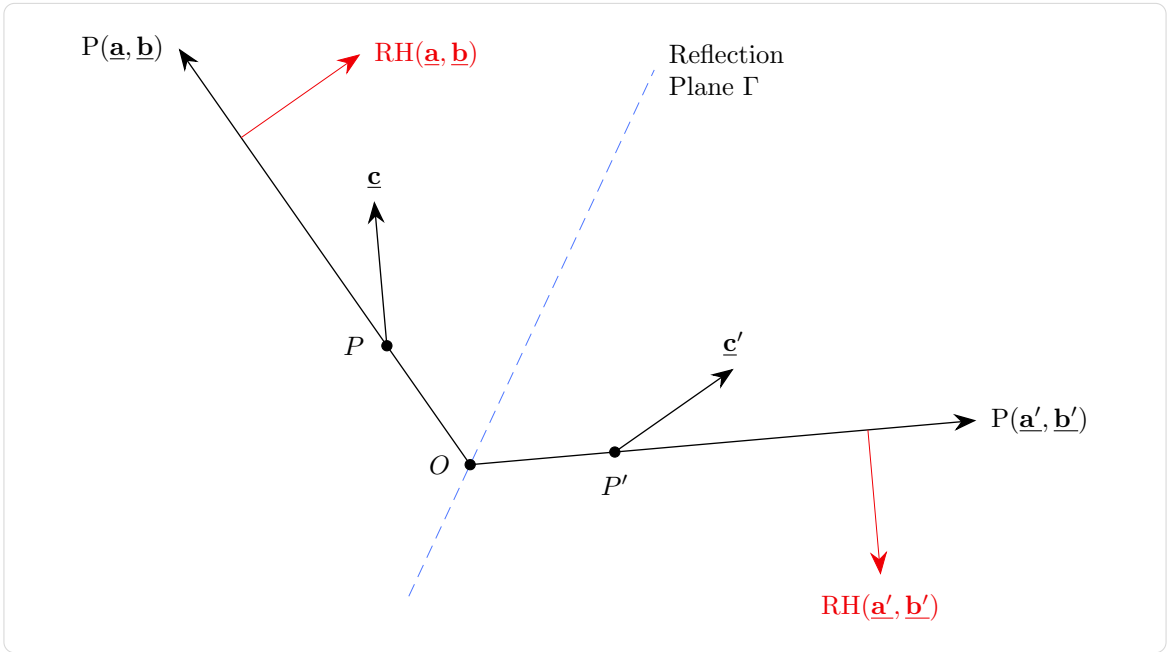
In Theorem 4 below we show that when an ordered list of three vectors is reflected in a plane then their handedness is flipped, ie. the 'mirror image' of the RH system is a LH system and vice-versa. Thus a system of three vectors has 'chirality' [18], ie. it cannot be overlayed onto its mirror image purely by rotations and translations alone, similarly to our left and right hands. Rotations and translations themselves clearly do not alter the handedness of a vector system.

Theorem 4 (Chirality Of System Of Three Vectors). If non-zero non-coplanar vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ form a RH system then these form a LH system after reflection in a plane, and vice-versa.

Proof. Consider a RH system $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ located at point P as shown in Figure 9, where the reflection plane Γ and the plane $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ are viewed perpendicularly to the page along their line of intersection at O. $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ lie in the plane $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ with tails at P in such a way that $\mathrm{RH}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ is above $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ in the diagram. Vector $\underline{\mathbf{c}}$ is shown at an arbitrary orientation (not necessarily in plane of page) but it must point to the side $\mathrm{RH}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ of $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ since $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ is a RH system. In the reflection, $\mathrm{RH}(\underline{\mathbf{a}}', \underline{\mathbf{b}}')$ must point to the downwards side of $\mathrm{P}(\underline{\mathbf{a}}', \underline{\mathbf{b}}')$, and this applies whichever way $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ are oriented within $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$. Thus since $\underline{\mathbf{c}}$ still points to the upwards side after reflection (ie. $\underline{\mathbf{c}}'$ points to the upward side of $\mathrm{P}(\underline{\mathbf{a}}', \underline{\mathbf{b}}')$) it follows that $\underline{\mathbf{a}}', \underline{\mathbf{b}}', \underline{\mathbf{c}}'$ is a LH system. The case of reflecting a LH system to obtain a RH system follows similarly by orienting $\underline{\mathbf{c}}$ to point below $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$.

QED

Figure 9: Chirality Of System Of Vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$

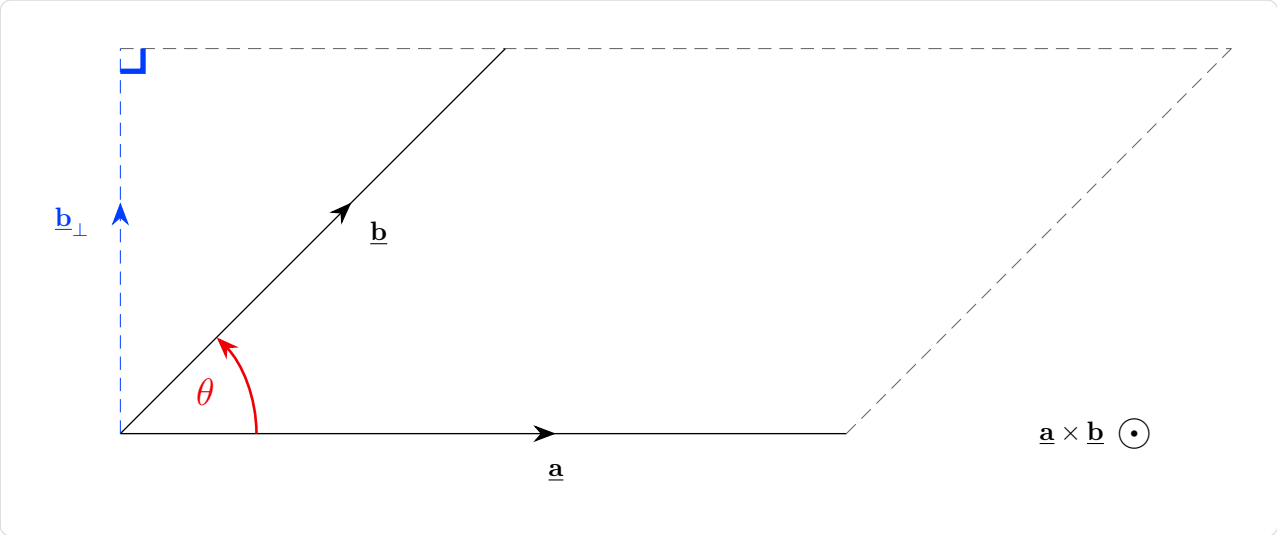


3.4 Vector Product In EV

The vector product $\underline{\mathbf{a}} \times \underline{\mathbf{b}}$ (aka cross product) of vectors $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ is a vector perpendicular to both $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$, and hence perpendicular to the plane contained by $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$. It is defined to have magnitude equal to the area $|\underline{\mathbf{a}}|\,|\underline{\mathbf{b}}|\sin\theta$ of the parallelogram defined by $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$, as shown in Figure 10, where θ is the unsigned minor angle between $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$. The direction of $\underline{\mathbf{a}} \times \underline{\mathbf{b}}$ is defined to be perpendicular to plane $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ and pointing towards $\mathrm{RH}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ (out of the page in Figure 10). Unlike in the case of scalar product we could place the vectors head to tail and use the minor angle $180° - \theta$, since the same parallelogram results and its area is $\sin(180° - \theta) = \sin\theta$. If $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ are parallel then $\underline{\mathbf{a}} \times \underline{\mathbf{b}}$ is defined to be zero since in this case there is no way from $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ alone to specify a vector perpendicular to both $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$, since there is an entire plane in EV comprising such vectors, so $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ do not define a unique plane in this case.

It is clear from Figure 10 that $\underline{\mathbf{a}} \times \underline{\mathbf{b}} = \underline{\mathbf{a}} \times \underline{\mathbf{b}}_\perp$, where $\underline{\mathbf{b}}_\perp$ is the component of $\underline{\mathbf{b}}$ perpendicular to $\underline{\mathbf{a}}$, since the direction will be the same (ie. out of the page), for any minor angle $\theta \in (0^\circ, 180^\circ)$, and the rectangle produced by $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}_\perp$ sits on the same base and has the same height as the original parallelogram so that it has the same area. We find below also that $\underline{\mathbf{a}} \times \underline{\mathbf{b}} = \underline{\mathbf{a}}_\perp \times \underline{\mathbf{b}}$, so it is only the perpendicular component of the vector that we are multiplying by that matters, the parallel component does not affect the result (in contrast to the scalar product (§3.2) where only the parallel component affects the result). Likewise only the perpendicular component contributes to defining the plane $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ — if there is no perpendicular component at all, ie. $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ are parallel, then $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ is undefined. The above is why it is natural to define the magnitude of $\underline{\mathbf{a}} \times \underline{\mathbf{b}}$ to be the area of the parallelogram in Figure 10, because this depends on the perpendicular component only.

Figure 10: Vector Product Definition



Vector product has the following properties :

1. $\underline{\mathbf{b}} \times \underline{\mathbf{a}} = -\underline{\mathbf{a}} \times \underline{\mathbf{b}}$ (anti-commutative)
2. $\underline{\mathbf{a}} \times \underline{\mathbf{b}} = \underline{\mathbf{a}} \times \underline{\mathbf{b}}_\perp = \underline{\mathbf{a}}_\perp \times \underline{\mathbf{b}}$, where $\underline{\mathbf{b}}_\perp$ is the component of $\underline{\mathbf{b}}$ perpendicular to $\underline{\mathbf{a}}$, and $\underline{\mathbf{a}}_\perp$ is the component of $\underline{\mathbf{a}}$ perpendicular to $\underline{\mathbf{b}}$
3. $(\lambda\underline{\mathbf{a}}) \times \underline{\mathbf{b}} = \underline{\mathbf{a}} \times (\lambda\underline{\mathbf{b}}) = \lambda(\underline{\mathbf{a}} \times \underline{\mathbf{b}})$ (multiplicative)
4. for non-zero $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$, $\underline{\mathbf{a}} \times \underline{\mathbf{b}} = \underline{\mathbf{0}} \iff \underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ are parallel
5. $(\underline{\mathbf{a}} + \underline{\mathbf{b}}) \times \underline{\mathbf{c}} = (\underline{\mathbf{a}} \times \underline{\mathbf{c}}) + (\underline{\mathbf{b}} \times \underline{\mathbf{c}})$ (distributive)

For the anti-commutative property we use the fact observed in §3.3 that $\mathrm{RH}(\underline{\mathbf{b}}, \underline{\mathbf{a}})$ is the opposite side of $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ from $\mathrm{RH}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$, whilst the parallelogram involved is the same. The first equality in property 2 was observed above, and the second follows from that by applying the anti-commutative property : $\underline{\mathbf{a}}_\perp \times \underline{\mathbf{b}} = -\underline{\mathbf{b}} \times \underline{\mathbf{a}}_\perp = -\underline{\mathbf{b}} \times \underline{\mathbf{a}} = \underline{\mathbf{a}} \times \underline{\mathbf{b}}$. The multiplicative property is clear — for the case $\lambda = -1$ the same parallelogram would be involved but the direction determined by the right hand rule would flip.

For the distributive property we note it suffices to prove it for the special case where both $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ are perpendicular to $\underline{\mathbf{c}}$, since then :

$$\begin{aligned}
(\underline{\mathbf{a}}+\underline{\mathbf{b}})\times\underline{\mathbf{c}} &= (\underline{\mathbf{a}}+\underline{\mathbf{b}})_{\perp}\times\underline{\mathbf{c}}\,, \quad \text{by property 2}\\
&= (\underline{\mathbf{a}}_{\perp}+\underline{\mathbf{b}}_{\perp})\times\underline{\mathbf{c}}\,, \quad \text{perpendicular of sum is sum of perpendiculars}\\
&= (\underline{\mathbf{a}}_{\perp}\times\underline{\mathbf{c}})+(\underline{\mathbf{b}}_{\perp}\times\underline{\mathbf{c}})\,, \quad \text{by the special case}\\
&= (\underline{\mathbf{a}}\times\underline{\mathbf{c}})+(\underline{\mathbf{b}}\times\underline{\mathbf{c}})\,, \quad \text{by property 2}
\end{aligned}$$

To prove the special case consider Figure 11. By hypothesis, $\underline{\mathbf{c}}$ is perpendicular to the plane containing $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$. (Note the case where $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ are parallel follows immediately from the multiplicative property above). In Figure 11, $\underline{\mathbf{c}}$ is directed into the page, so that by the right-hand rule the three vector products are directed as shown at right angles to the three sides of the triangle, and we have $|\underline{\mathbf{a}}\times\underline{\mathbf{c}}| = |\underline{\mathbf{a}}|\,|\underline{\mathbf{c}}|, |\underline{\mathbf{b}}\times\underline{\mathbf{c}}| = |\underline{\mathbf{b}}|\,|\underline{\mathbf{c}}|, |(\underline{\mathbf{a}}+\underline{\mathbf{b}})\times\underline{\mathbf{c}}| = |\underline{\mathbf{a}}+\underline{\mathbf{b}}|\,|\underline{\mathbf{c}}|$. Thus the lengths of the vectors $(\underline{\mathbf{a}}\times\underline{\mathbf{c}})/|\underline{\mathbf{c}}|, (\underline{\mathbf{b}}\times\underline{\mathbf{c}})/|\underline{\mathbf{c}}|, ((\underline{\mathbf{a}}+\underline{\mathbf{b}})\times\underline{\mathbf{c}})/|\underline{\mathbf{c}}|$ are equal to the lengths of the corresponding sides $\underline{\mathbf{a}}, \underline{\mathbf{b}}, (\underline{\mathbf{a}}+\underline{\mathbf{b}})$ of the triangle to which they are perpendicular. But observing the orientations of these three vectors $(\underline{\mathbf{a}}\times\underline{\mathbf{c}})/|\underline{\mathbf{c}}|, (\underline{\mathbf{b}}\times\underline{\mathbf{c}})/|\underline{\mathbf{c}}|, ((\underline{\mathbf{a}}+\underline{\mathbf{b}})\times\underline{\mathbf{c}})/|\underline{\mathbf{c}}|$ we note each is a $+90^{\circ}$ (ie. ACW) rotation of its corresponding side $\underline{\mathbf{a}}, \underline{\mathbf{b}}, (\underline{\mathbf{a}}+\underline{\mathbf{b}})$ of the triangle. This means the vectors $(\underline{\mathbf{a}}\times\underline{\mathbf{c}})/|\underline{\mathbf{c}}|, (\underline{\mathbf{b}}\times\underline{\mathbf{c}})/|\underline{\mathbf{c}}|, ((\underline{\mathbf{a}}+\underline{\mathbf{b}})\times\underline{\mathbf{c}})/|\underline{\mathbf{c}}|$ are exactly equal to the sides of the original triangle after it has undergone a rotation in the plane (about any point) of $+90^{\circ}$. Thus from this rotated triangle we obtain :

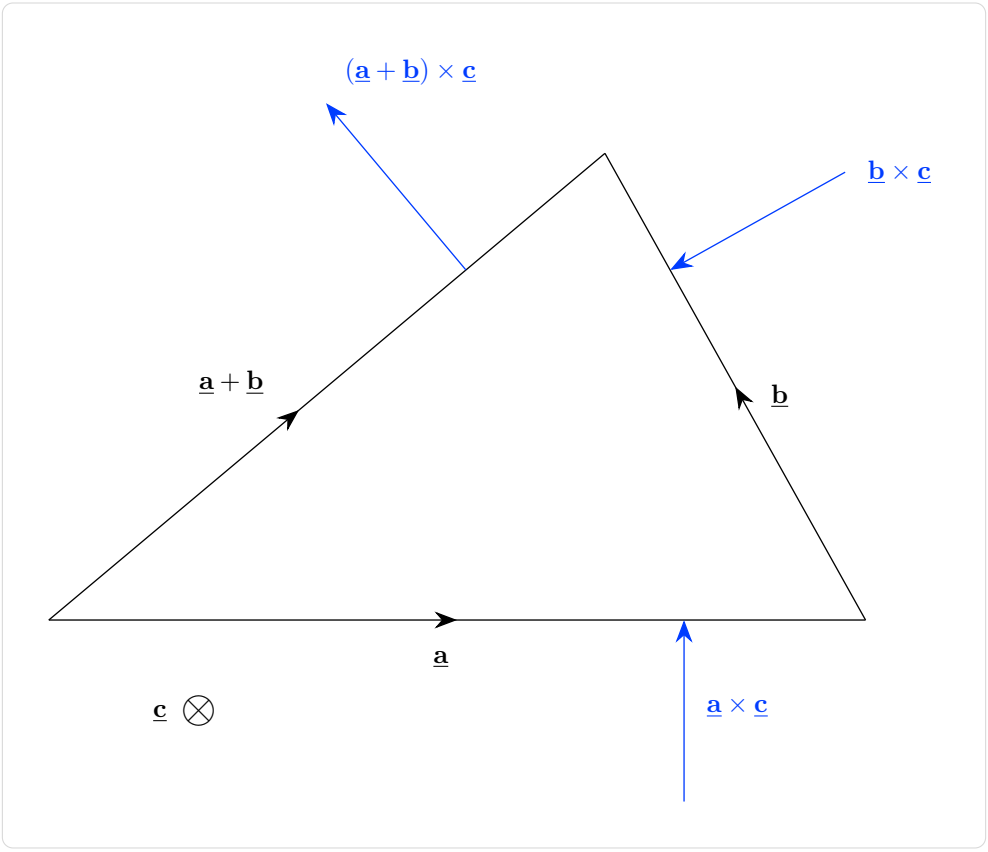
$$\frac{1}{|\underline{\mathbf{c}}|}\left((\underline{\mathbf{a}}+\underline{\mathbf{b}})\times\underline{\mathbf{c}}\right) = \frac{1}{|\underline{\mathbf{c}}|}(\underline{\mathbf{a}}\times\underline{\mathbf{c}}) + \frac{1}{|\underline{\mathbf{c}}|}(\underline{\mathbf{b}}\times\underline{\mathbf{c}})$$

and hence

$$(\underline{\mathbf{a}}+\underline{\mathbf{b}})\times\underline{\mathbf{c}} = (\underline{\mathbf{a}}\times\underline{\mathbf{c}})+(\underline{\mathbf{b}}\times\underline{\mathbf{c}})$$

as required. The case where $\underline{\mathbf{c}}$ points out of the page follows similarly, the only difference being the triangle undergoes a -90° rotation (ie. CW).

Figure 11: Vector Product Distributive Law



3.5 Scalar Triple Product In EV

Definition 4. The scalar triple product of an ordered list of three vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, and $\underline{\mathbf{c}}$ in EV is defined to be the real number $\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}})$.

□

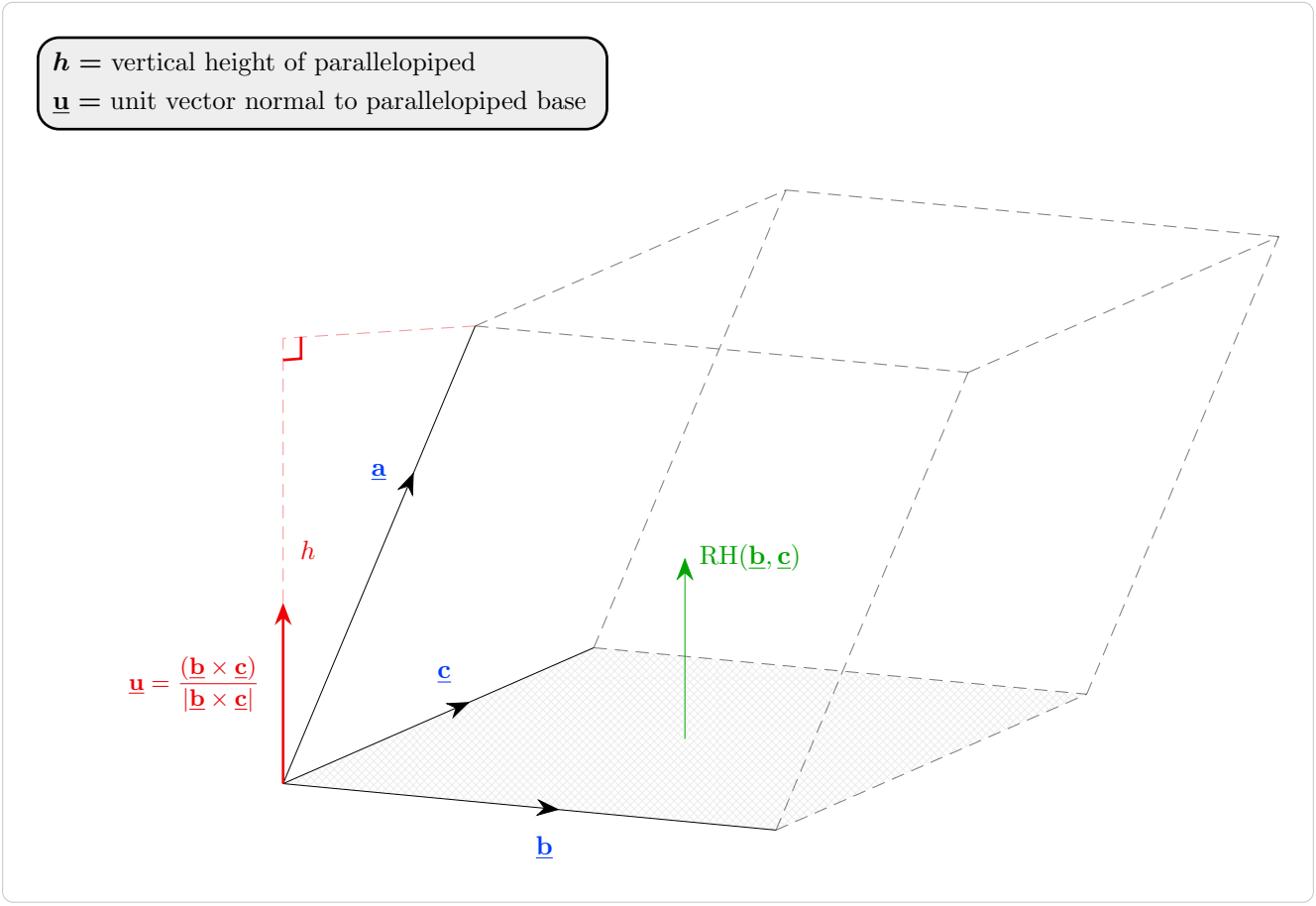
Definition 5. $V(\underline{\mathbf{a}}, \underline{\mathbf{b}}, \underline{\mathbf{c}})$ is the (unsigned) volume of the parallelopiped (Figure 12) defined by vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$. The value of $V(\underline{\mathbf{a}}, \underline{\mathbf{b}}, \underline{\mathbf{c}})$ is taken to be zero if any one of $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, or $\underline{\mathbf{c}}$ is zero, or if $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ are coplanar.

□

Figure 12 shows the parallelopiped defined by a RH system $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ (note by Theorem 3 this implies $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$, $\underline{\mathbf{a}}$ is a RH system).

A parallelopiped is a particular type of oblique prism which has a parallelogram as its base. Three vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ define the same parallelopiped whichever of the six possible reorderings of $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ is chosen – we take any two of the vectors to form the parallelogram base and the remaining vector as the prism base translation vector. Since we have defined $V(\underline{\mathbf{a}}, \underline{\mathbf{b}}, \underline{\mathbf{c}})$ to be an unsigned volume it has the same non-negative value for both RH and LH reorderings of $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$, since both RH and LH reorderings result in the same parallelopiped, ie. although handedness and scalar triple product depend on the ordering of the three vectors the parallelopiped does not. In Theorem 5 below the order dependent scalar triple product provides a 'signed volume' for the parallelopiped, which is positive for RH systems and negative for LH systems and zero when the figure becomes flat or a line or point.

Figure 12: Parallelopiped Defined By RH Vector System $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$



Theorem 5. The scalar triple product $\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}})$ in EV is given by :

$$\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) = \begin{cases} V(\underline{\mathbf{a}}, \underline{\mathbf{b}}, \underline{\mathbf{c}}), & \text{if } \underline{\mathbf{a}}, \underline{\mathbf{b}}, \underline{\mathbf{c}} \text{ is a RH system,} \\ -V(\underline{\mathbf{a}}, \underline{\mathbf{b}}, \underline{\mathbf{c}}), & \text{if } \underline{\mathbf{a}}, \underline{\mathbf{b}}, \underline{\mathbf{c}} \text{ is a LH system,} \\ 0, & \text{if any one of } \underline{\mathbf{a}}, \underline{\mathbf{b}}, \underline{\mathbf{c}} \text{ is zero, or these three vectors are coplanar.} \end{cases}$$

Proof. The result is clear if any one of $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ is zero or if $\underline{\mathbf{b}}$ and $\underline{\mathbf{c}}$ are parallel. If $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ are coplanar then $\underline{\mathbf{a}}$ lies in the plane $\mathrm{P}(\underline{\mathbf{b}}, \underline{\mathbf{c}})$ and hence is perpendicular to $\underline{\mathbf{b}} \times \underline{\mathbf{c}}$, so the scalar triple product is zero.

Consider a general RH system $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$. From Theorem 3, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$, $\underline{\mathbf{a}}$ must also be a RH system so that $\underline{\mathbf{a}}$ points to the side $\mathrm{RH}(\underline{\mathbf{b}}, \underline{\mathbf{c}})$ of plane $\mathrm{P}(\underline{\mathbf{b}}, \underline{\mathbf{c}})$, as shown in Figure 12. Generally the volume of a prism is the area of the base times the vertical height, so we have :

$$\begin{aligned} V(\underline{\mathbf{a}}, \underline{\mathbf{b}}, \underline{\mathbf{c}}) &= h \cdot |\underline{\mathbf{b}} \times \underline{\mathbf{c}}|, \quad \text{from the definition of vector product (§3.4)} \\ &= (\underline{\mathbf{a}} \cdot \underline{\mathbf{u}}) \cdot |\underline{\mathbf{b}} \times \underline{\mathbf{c}}|, \quad \text{since again from the definition of vector product (§3.4), } \underline{\mathbf{a}} \text{ and } \underline{\mathbf{u}} \text{ lie to} \\ &\quad \text{the same side } \mathrm{RH}(\underline{\mathbf{b}}, \underline{\mathbf{c}}) \text{ of plane } \mathrm{P}(\underline{\mathbf{b}}, \underline{\mathbf{c}}) \\ &= \underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) \end{aligned}$$

Consider a general LH system $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$. From Theorem 3, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$, $\underline{\mathbf{a}}$ must also be a LH system so that $\underline{\mathbf{a}}$ points to the side $\mathrm{LH}(\underline{\mathbf{b}}, \underline{\mathbf{c}})$ of plane $\mathrm{P}(\underline{\mathbf{b}}, \underline{\mathbf{c}})$, ie. to the opposite side shown in Figure 12. The vector $\underline{\mathbf{u}}$ still points to the upper side in Figure 12 so $\underline{\mathbf{a}}$ and $-\underline{\mathbf{u}}$ lie to the same side $\mathrm{LH}(\underline{\mathbf{b}}, \underline{\mathbf{c}})$ of plane $\mathrm{P}(\underline{\mathbf{b}}, \underline{\mathbf{c}})$ so $h = \underline{\mathbf{a}} \cdot (-\underline{\mathbf{u}})$ and thus :

$$V(\underline{\mathbf{a}}, \underline{\mathbf{b}}, \underline{\mathbf{c}}) = -\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}})$$

QED

Theorem 6. In EV the value of the scalar triple product $\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}})$ is unchanged after any cyclic reordering of the three vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$, ie.

$$\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) = \underline{\mathbf{c}} \cdot (\underline{\mathbf{a}} \times \underline{\mathbf{b}}) = \underline{\mathbf{b}} \cdot (\underline{\mathbf{c}} \times \underline{\mathbf{a}}).$$

Interchanging any two of the vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ in $\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}})$ flips the sign of the scalar triple product, ie.

$$\underline{\mathbf{b}} \cdot (\underline{\mathbf{a}} \times \underline{\mathbf{c}}) = \underline{\mathbf{a}} \cdot (\underline{\mathbf{c}} \times \underline{\mathbf{b}}) = \underline{\mathbf{c}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{a}}) = -\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}).$$

Proof. First of all consider the case where one of $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ is zero, or these three vectors are coplanar. By Theorem 5 this implies that the six scalar triple products that come from the six possible reorderings of $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ are all zero and hence the required result follows.

Consider the case of a RH system $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$. By Theorem 3 both $\underline{\mathbf{c}}$, $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$ and $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$, $\underline{\mathbf{a}}$ are RH systems and so from Theorem 5 we have :

$$\begin{aligned} \underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) &= V(\underline{\mathbf{a}}, \underline{\mathbf{b}}, \underline{\mathbf{c}}) \\ \underline{\mathbf{c}} \cdot (\underline{\mathbf{a}} \times \underline{\mathbf{b}}) &= V(\underline{\mathbf{c}}, \underline{\mathbf{a}}, \underline{\mathbf{b}}) \\ \underline{\mathbf{b}} \cdot (\underline{\mathbf{c}} \times \underline{\mathbf{a}}) &= V(\underline{\mathbf{b}}, \underline{\mathbf{c}}, \underline{\mathbf{a}}) \end{aligned}$$

But the three volumes on the RHS's here are all the unsigned volume of the same parallelopiped figure defined by $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ — as we mentioned above this quantity does not depend on the order of $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$. Thus the first three scalar triple products, obtained by cycling, are equal, as required.

In §3.3 we observed that interchanging any two of the non-zero non-coplanar vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ flips the handedness. Thus each of the triples $\underline{\mathbf{b}}$, $\underline{\mathbf{a}}$, $\underline{\mathbf{c}}$ and $\underline{\mathbf{a}}$, $\underline{\mathbf{c}}$, $\underline{\mathbf{b}}$ and $\underline{\mathbf{c}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{a}}$ are all LH (since $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ is RH). Then from Theorem 5 we have :

$$\begin{aligned} \underline{\mathbf{b}} \cdot (\underline{\mathbf{a}} \times \underline{\mathbf{c}}) &= -V(\underline{\mathbf{b}}, \underline{\mathbf{a}}, \underline{\mathbf{c}}) \\ \underline{\mathbf{a}} \cdot (\underline{\mathbf{c}} \times \underline{\mathbf{b}}) &= -V(\underline{\mathbf{a}}, \underline{\mathbf{c}}, \underline{\mathbf{b}}) \\ \underline{\mathbf{c}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{a}}) &= -V(\underline{\mathbf{c}}, \underline{\mathbf{b}}, \underline{\mathbf{a}}) \end{aligned}$$

the three volumes on the RHS again being the unsigned volume of the same parallelopiped figure defined by $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$. Thus the second group of three scalar triple products are equal with value minus the common value of the first group. Alternatively we could see the latter using the anti-commutative property of the vector product.

The case of a LH system $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ follows similarly.

QED

3.6 Vector Triple Product In EV

Definition 6. The vector triple product of an ordered list of three vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, and $\underline{\mathbf{c}}$ in EV is defined to be the vector $\underline{\mathbf{a}} \times (\underline{\mathbf{b}} \times \underline{\mathbf{c}})$.

□

Theorem 7. For any $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ in EV, $\underline{\mathbf{a}} \times (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) = (\underline{\mathbf{a}} \cdot \underline{\mathbf{c}})\,\underline{\mathbf{b}} - (\underline{\mathbf{a}} \cdot \underline{\mathbf{b}})\,\underline{\mathbf{c}}$.

Proof. The cases of $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, or $\underline{\mathbf{c}}$ zero or $\underline{\mathbf{b}}$ parallel to $\underline{\mathbf{c}}$ are clear. Otherwise we first note since $\underline{\mathbf{r}} = \underline{\mathbf{a}} \times (\underline{\mathbf{b}} \times \underline{\mathbf{c}})$ is perpendicular to $\underline{\mathbf{b}} \times \underline{\mathbf{c}}$ which is in turn perpendicular to the plane $\mathrm{P}(\underline{\mathbf{b}}, \underline{\mathbf{c}})$ containing $\underline{\mathbf{b}}$ and $\underline{\mathbf{c}}$, so $\underline{\mathbf{r}}$ must also lie within this plane, and hence $\underline{\mathbf{r}} = \lambda\,\underline{\mathbf{b}} + \mu\,\underline{\mathbf{c}}$ for $\lambda, \mu \in \mathbb{R}$ as given by Theorem 2 applied to $\mathrm{P}(\underline{\mathbf{b}}, \underline{\mathbf{c}})$. Using the notation of Theorem 2 we can take normals to $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ lying within plane $\mathrm{P}(\underline{\mathbf{b}}, \underline{\mathbf{c}})$ to be $\underline{\mathbf{m}} = (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) \times \underline{\mathbf{b}}$ and $\underline{\mathbf{n}} = (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) \times \underline{\mathbf{c}}$ respectively, as shown in Figure 13. Then

$$\lambda = \frac{(\underline{\mathbf{r}} \cdot \underline{\mathbf{n}})}{(\underline{\mathbf{b}} \cdot \underline{\mathbf{n}})}, \qquad \mu = \frac{(\underline{\mathbf{r}} \cdot \underline{\mathbf{m}})}{(\underline{\mathbf{c}} \cdot \underline{\mathbf{m}})}.$$

From Figure 13 we have :

$$\underline{\mathbf{b}} \cdot \underline{\mathbf{n}} = |\underline{\mathbf{b}}|\,|\underline{\mathbf{n}}| \cos(\theta + 90^\circ) = -|\underline{\mathbf{b}}|\,|\underline{\mathbf{c}}|\,|\underline{\mathbf{b}} \times \underline{\mathbf{c}}| \sin\theta = -|\underline{\mathbf{b}} \times \underline{\mathbf{c}}|^2$$

and by Theorem 6, cycling the scalar triple product :

$$\underline{\mathbf{r}} \cdot \underline{\mathbf{n}} = \underline{\mathbf{n}} \cdot \underline{\mathbf{r}} = \underline{\mathbf{n}} \cdot (\underline{\mathbf{a}} \times (\underline{\mathbf{b}} \times \underline{\mathbf{c}})) = \underline{\mathbf{a}} \cdot ((\underline{\mathbf{b}} \times \underline{\mathbf{c}}) \times \underline{\mathbf{n}})$$

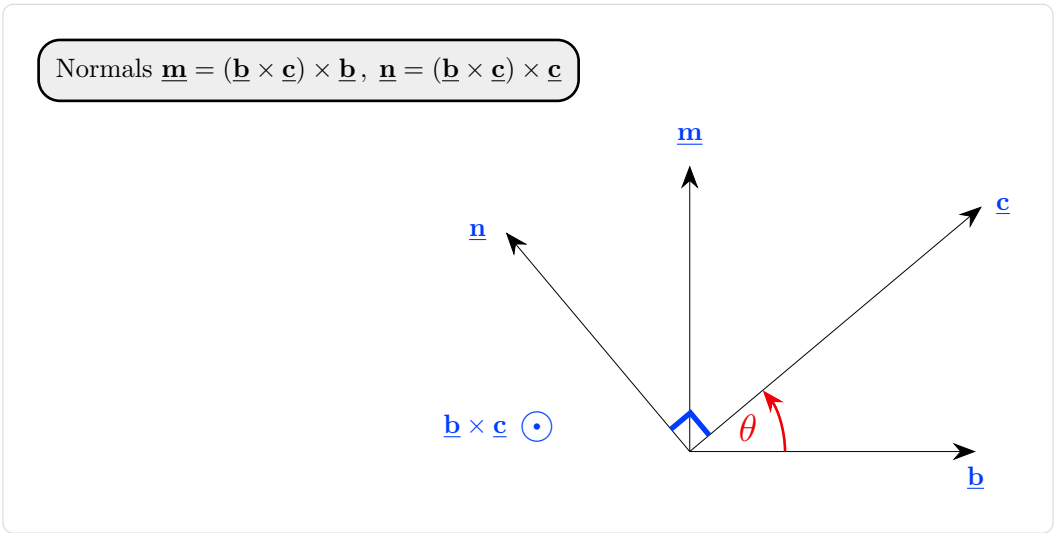
where using $|\underline{\mathbf{n}}| = |\underline{\mathbf{b}} \times \underline{\mathbf{c}}|\,|\underline{\mathbf{c}}|$ we have :

$$(\underline{\mathbf{b}} \times \underline{\mathbf{c}}) \times \underline{\mathbf{n}} = |\underline{\mathbf{b}} \times \underline{\mathbf{c}}|\,|\underline{\mathbf{n}}| \cdot \left(-\frac{\underline{\mathbf{c}}}{|\underline{\mathbf{c}}|}\right) = -|\underline{\mathbf{b}} \times \underline{\mathbf{c}}|^2\,\underline{\mathbf{c}}$$

and thus $\lambda = \underline{\mathbf{a}} \cdot \underline{\mathbf{c}}$, as required. A similar argument using the normal vector $\underline{\mathbf{m}}$ gives $\mu = -\underline{\mathbf{a}} \cdot \underline{\mathbf{b}}$.

QED

Figure 13: Vector Triple Product $\underline{\mathbf{a}} \times (\underline{\mathbf{b}} \times \underline{\mathbf{c}})$ In Plane $\mathrm{P}(\underline{\mathbf{b}}, \underline{\mathbf{c}})$



4 Coordinate Systems

4.1 Cartesian Coordinate System

A Cartesian coordinate system (CCS) [19] for ES consists of 3 mutually perpendicular axes, x, y, and z. If the positive x, y, and z axes, as vectors, form a right-handed system (according to Definition 3) then we call it a RH CCS and if these vectors form a left-handed system we call it a LH CCS. The CCS can be oriented in any manner within ES.

The two things we can do with a coordinate system (whether Cartesian or other type) are (i) take an equation or inequation in x, y, and z and plot the region of ES that satisfies the equation, and (ii) take a region of ES determined geometrically and then ask what the associated coordinates are wrt the coordinate system.

Every point P in ES defines a coordinate cuboid with interior diagonal $\overrightarrow{OP}$ and faces parallel to the yz, xz, and xy coordinate planes as shown in Figure 14. The x, y, z coordinates of P are then defined to be the points on the x, y, z axes that intersect with the coordinate cuboid, and these may be positive or negative. The coordinate cuboid becomes a 2D rectangle, a line, or a point respectively as one, two, or three of the three coordinates are zero. The coordinates of every point are unique so this gives a one-to-one mapping from ES to $\mathbb{R}^3$. The coordinates (x, y, z) form an element of $\mathbb{R}^3$ called the 'coordinate vector' of P wrt CCS, [20].

Every vector $\underline{\mathbf{r}}$ in EV can be associated with the DLS $\overrightarrow{OP}$ for a unique point P in ES. Thus $\underline{\mathbf{r}}$ is associated with a unique set of coordinates $(x, y, z) \in \mathbb{R}^3$, which is then the coordinate vector of $\underline{\mathbf{r}}$ wrt CCS. Because vectors only depend on length and direction these coordinates are the same even after any pure translation of CCS within ES (ie. no rotation involved), ie. the placement of the origin O of CCS does not affect the vector coordinates. The coordinates of *points* within ES however are changed by the translation of CCS.

If we let $\underline{\mathbf{u}}$, $\underline{\mathbf{v}}$, $\underline{\mathbf{w}}$ be unit vectors along the positive x, y, and z axes[1] then it is clear that :

$$\underline{\mathbf{r}} = x\,\underline{\mathbf{u}} + y\,\underline{\mathbf{v}} + z\,\underline{\mathbf{w}}$$

from the definition of addition and scalar multiplication on the vector space EV. The values x, y, and z satisfying this identity must be unique since every distinct ordered triple (x, y, z) produces a unique point P and hence unique DLS $\overrightarrow{OP}$. In the language of Linear Algebra this means that $\underline{\mathbf{u}}$, $\underline{\mathbf{v}}$, $\underline{\mathbf{w}}$ form a basis for the vector space EV [10, p112], and x, y, z are the unique coefficients of vector $\underline{\mathbf{r}}$ wrt that basis. A basis for a general vector space is a set of vectors which is linearly independent (LI) [10, p106] and spans the space [10, p102], and this is equivalent to saying that every vector in the space is a unique linear combination of the basis vectors, as we have just shown for the vectors $\underline{\mathbf{u}}$, $\underline{\mathbf{v}}$, and $\underline{\mathbf{w}}$ in the vector space EV.

The unit vectors $\underline{\mathbf{u}}$, $\underline{\mathbf{v}}$, $\underline{\mathbf{w}}$ are the 'standard basis' for CCS (in a similar fashion to [10, p113]). The coefficients (x, y, z) wrt the standard basis are the same as the coordinates of $\underline{\mathbf{r}} = \overrightarrow{OP}$ wrt CCS. Note there is not a concept of a 'standard basis' for EV itself considered on its own, as there is no reference system against which to define it, however we can define a standard basis for EV wrt any given CCS that is placed within ES. The vector space $\mathbb{R}^3$ has as standard basis $(1, 0, 0), (0, 1, 0), (0, 0, 1)$.

Instead of unit vectors we could take any non-zero vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ directed along the respective positive axes so that $\underline{\mathbf{u}}$, $\underline{\mathbf{v}}$, $\underline{\mathbf{w}}$ are fixed positive scalar multiples of these. Then any vector $\underline{\mathbf{r}}$ could be expressed as a linear combination in EV of form :

$$\underline{\mathbf{r}} = x'\,\underline{\mathbf{a}} + y'\,\underline{\mathbf{b}} + z'\,\underline{\mathbf{c}}$$

where the scalar values x', y', z' are unique due to the uniqueness of x, y, z above. Then $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ also form a basis for EV and $(x', y', z') \in \mathbb{R}^3$ is the coordinate vector of $\underline{\mathbf{r}}$ wrt $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ (the term 'coefficents' wrt $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ is also used for x', y', z'). These are non-standard basis vectors for CCS and they amount to different units of measurement in the x, y, and z directions. But ES only has a single unit of measurement applying equally in all directions (ie. it is 'isotropic'), so the coordinates (x', y', z') of vector $\underline{\mathbf{r}}$ wrt CCS with the non-standard basis vectors are not coordinates of point P in ES, as they were in the case of the standard basis for CCS. Plotting an equation in (x', y', z') coordinates in ES would in general result in a scaled figure with a distorted shape, for example plotting the sphere equation in (x', y', z') coordinates would result in an ellipsoid in ES. Normally a CCS uses the standard basis vectors $\underline{\mathbf{u}}$, $\underline{\mathbf{v}}$, $\underline{\mathbf{w}}$, [19], however considering the non-standard basis vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ leads us into considering the general 'skew' coordinate system in §4.4.

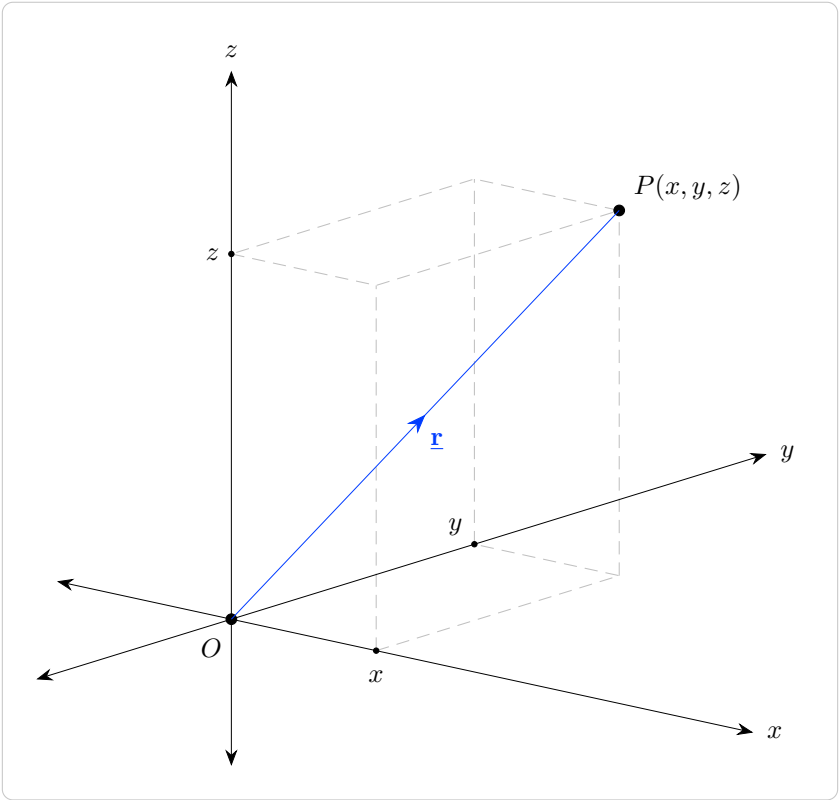
For any non-standard basis $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ of CCS we thus have a bijective mapping from the vector space EV to the vector space $\mathbb{R}^3$ defined by taking the coordinate vector (x', y', z') of $\underline{\mathbf{r}} \in$ EV wrt $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$. It can be readily checked that this is a

[1] To avoid confusion we use the notation $\underline{\mathbf{u}}$, $\underline{\mathbf{v}}$, $\underline{\mathbf{w}}$ rather than $\underline{\mathbf{i}}$, $\underline{\mathbf{j}}$, $\underline{\mathbf{k}}$ as we will use $\underline{\mathbf{i}}$, $\underline{\mathbf{j}}$, $\underline{\mathbf{k}}$ later to denote the unit vectors (1, 0, 0), (0, 1, 0), and (0, 0, 1) within $\mathbb{R}^3$.

linear mapping [10, p133], and hence a vector space isomorphism from EV to $\mathbb{R}^3$ [10, p178]. Thus EV and $\mathbb{R}^3$ are one and the same as vector spaces, but the objects they comprise are quite different — an element of EV is an ordered pair (l, d) consisting of a real number l representing a length together with a direction $d \in \mathbb{D}$ as a primitive notion (§2), whereas an element of $\mathbb{R}^3$ is an ordered triple of real numbers. It is a category error to conflate these two element types as being the same thing, though in practice we would often refer to one of the types in terms of the other with the distinction implicitly understood. For example if $\underline{\mathbf{d}}, \underline{\mathbf{e}}, \underline{\mathbf{f}} \in$ EV and if a matrix was denoted by $[\ \underline{\mathbf{d}} \quad \underline{\mathbf{e}} \quad \underline{\mathbf{f}}\ ]$ then the columns of the matrix would be understood to be the $\mathbb{R}^3$ representations of $\underline{\mathbf{d}}, \underline{\mathbf{e}}, \underline{\mathbf{f}}$ wrt some given CCS.

Note that throughout the above the CCS can be either RH or LH.

Figure 14: RH Cartesian Coordinate System



4.1.1 EV Scalar Product Expressed In Cartesian Coordinates

In the following the CCS can be either RH or LH. Note in this section we are asking what the geometric EV scalar product would look like in terms of a CCS, we are not at this point formally defining scalar product on $\mathbb{R}^3$ (see §4.2).

If vectors $\underline{\mathbf{a}}, \underline{\mathbf{b}} \in$ EV have coordinate vectors $(a_1, a_2, a_3), (b_1, b_2, b_3) \in \mathbb{R}^3$ wrt the standard basis vectors $\underline{\mathbf{u}}, \underline{\mathbf{v}}, \underline{\mathbf{w}} \in$ EV of CCS then from the properties 1–4 of scalar product in §3.2 we have :

$$\begin{aligned} & \underline{\mathbf{a}} \cdot \underline{\mathbf{b}} \quad \text{in EV} \quad = \quad (a_1\,\underline{\mathbf{u}} + a_2\,\underline{\mathbf{v}} + a_3\,\underline{\mathbf{w}}) \cdot (b_1\,\underline{\mathbf{u}} + b_2\,\underline{\mathbf{v}} + b_3\,\underline{\mathbf{w}}) \\ \Rightarrow \quad & \underline{\mathbf{a}} \cdot \underline{\mathbf{b}} \quad \text{in EV} \quad = \quad a_1 b_1 + a_2 b_2 + a_3 b_3 \qquad \text{(RH or LH CCS)} \end{aligned} \tag{5}$$

since for the standard basis vectors we have $\underline{\mathbf{u}} \cdot \underline{\mathbf{u}} = \underline{\mathbf{v}} \cdot \underline{\mathbf{v}} = \underline{\mathbf{w}} \cdot \underline{\mathbf{w}} = 1$ and $\underline{\mathbf{u}} \cdot \underline{\mathbf{v}} = \underline{\mathbf{v}} \cdot \underline{\mathbf{w}} = \underline{\mathbf{u}} \cdot \underline{\mathbf{w}} = 0$.

Note this means that for a given $\underline{\mathbf{a}}, \underline{\mathbf{b}} \in$ EV the quantity $a_1 b_1 + a_2 b_2 + a_3 b_3$ must be the same for all possible orientations of the CCS in 3D space.

4.1.2 Determinant Notation

Although we have not yet defined the term 'determinant', which is a term applying to a general $n \times n$ matrix, for the following discussions on the vector product expressed in Cartesian coordinates it is convenient at this point to define the notation for the 2×2 and 3×3 determinants.

The form of the 3×3 determinant arises naturally when we express the vector product in terms of a CCS (§4.1.4) or when we try and solve for the vector at right angles to $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ using CCS coordinates (§4.1.3). The vector product, or the problem of solving for the perpendicular to a plane, when expressed in CCS coordinates, is the origin of the mathematical form of the determinant, and this form arises prior to any concept of a matrix.

The concepts of a matrix and of matrix multiplication arise when we consider the transform of coordinates from a CCS to a general skew CS (§4.5). The connection of the determinant (as defined in equation (7) below) to a matrix then arises when we use that transform to solve the matrix inverse problem in $\mathbb{R}^{3\times 3}$ (§7). When we formally define the matrix determinant $\det A$ in §7 and §10 later we back check it is consistent with the definition of the determinant notation in equation (7).

The determinant form is generalized to arbitrary $n \times n$ matrices over a general field $\mathbb{F}$ (§12.1) by conjecturing that the algebraic pattern we observe within the 3×3 case extends to the higher dimensions.

Definition 7 (2×2 Determinant). If a, b, c, d are real or complex numbers or elements of a field (or members of any algebraic structures with suitable operations) we define the 2×2 determinant by :

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \tag{6}$$

□

Definition 8 (3×3 Determinant). If $a, b, c, d, e, f, g, h, i$ are real or complex numbers or elements of a field (or members of any algebraic structures with suitable operations) we define the 3×3 determinant by :

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} \tag{7}$$

□

An important property of this notation is that the same value results if we 'transpose' the array of elements in the argument ie, if we change the rows into the columns (and hence the columns into the rows), or equivalently the element at position (r, c) goes to position (c, r), ie. it is reflected in the diagonal. This is clear in the 2×2 case. For the 3×3 case, by comparing terms we obtain :

$$\begin{vmatrix} a & d & g \\ b & e & h \\ c & f & i \end{vmatrix} = a \begin{vmatrix} e & h \\ f & i \end{vmatrix} - d \begin{vmatrix} b & h \\ c & i \end{vmatrix} + g \begin{vmatrix} b & e \\ c & f \end{vmatrix} = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} \tag{8}$$

Below we use the 3×3 determinant notation with EV or $\mathbb{R}^3$ vectors in the top line so it evaluates to a linear combination of vectors.

4.1.3 Alternative Approach To Vector Product Using Scalar Product And CCS

Above we defined the vector product on EV in such a way to produce a vector perpendicular to the plane contained by $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$, which obeyed the RH rule, and which depended only on the perpendicular component of the vector we multiplied by. This was done using a purely geometric method without using any coordinate system.

With a CCS (RH or LH) and the scalar product we could adopt an alternative approach to defining the vector product $\underline{\mathbf{a}} \times \underline{\mathbf{b}}$ which is purely algebraic by seeking a non-zero vector $\underline{\mathbf{c}} \in \mathrm{EV}$ with components $(c_1, c_2, c_3) \in \mathbb{R}^3$ wrt CCS satisfying $\underline{\mathbf{a}} \cdot \underline{\mathbf{c}} = 0$ and $\underline{\mathbf{b}} \cdot \underline{\mathbf{c}} = 0$. We would have to ensure the definition is consistent, ie. the EV vector produced is independent of which CCS we use. From §4.1.1 we require :

$$a_1 c_1 + a_2 c_2 + a_3 c_3 = 0 \tag{9}$$

$$b_1 c_1 + b_2 c_2 + b_3 c_3 = 0 \tag{10}$$

where $(a_1, a_2, a_3), (b_1, b_2, b_3) \in \mathbb{R}^3$ are the components of $\underline{\mathbf{a}}$, $\underline{\mathbf{b}} \in$ EV wrt CCS. Multiplying by b_1 and a_1 respectively we obtain :

$$\begin{aligned} b_1a_1c_1 + b_1a_2c_2 + b_1a_3c_3 &= 0 \quad (11) \\ a_1b_1c_1 + a_1b_2c_2 + a_1b_3c_3 &= 0 \quad (12) \end{aligned}$$

and then subtracting these :

$$(a_2b_1 - a_1b_2)c_2 + (a_3b_1 - a_1b_3)c_3 = 0$$

which is satisfied by :

$$\begin{aligned} c_2 &= a_3b_1 - a_1b_3 \\ c_3 &= a_1b_2 - a_2b_1 \end{aligned}$$

Then to satisfy (11) and (12) we require :

$$\begin{aligned} a_1b_1c_1 &= -a_2b_1c_2 - a_3b_1c_3 \\ &= -a_2b_1(a_3b_1 - a_1b_3) - a_3b_1(a_1b_2 - a_2b_1) \\ &= a_1b_1(a_2b_3 - a_3b_2) \end{aligned}$$

which is satisfied by :

$$c_1 = a_2b_3 - a_3b_2$$

If $b_1 \neq 0$ and $a_1 \neq 0$, from (11) and (12) we then obtain the required (9) and (10), and so in this case a solution to (9) and (10) is :

$$\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{vmatrix} \underline{\mathbf{i}} & \underline{\mathbf{j}} & \underline{\mathbf{k}} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \quad (13)$$

where $\underline{\mathbf{i}}$, $\underline{\mathbf{j}}$, $\underline{\mathbf{k}}$ denote the unit vectors (1, 0, 0), (0, 1, 0), and (0, 0, 1) in $\mathbb{R}^3$. It is readily checked that this is a solution in all other cases too. The vector $\underline{\mathbf{c}} = c_1\,\underline{\mathbf{u}} + c_2\,\underline{\mathbf{v}} + c_3\,\underline{\mathbf{w}}$ would then be a vector in EV perpendicular to the plane containing $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ (where $\underline{\mathbf{u}}$, $\underline{\mathbf{v}}$, $\underline{\mathbf{w}}$ are the standard basis vectors for CCS), as desired. We observe that the vector $-\underline{\mathbf{c}}$ would serve just as well.

To compare this with the geometric vector product firstly consider the magnitude of the vector $\underline{\mathbf{c}}$:

$$\begin{aligned} |\underline{\mathbf{c}}|^2 &= (a_2b_3 - a_3b_2)^2 + (a_1b_3 - a_3b_1)^2 + (a_1b_2 - a_2b_1)^2 \\ &= a_1^2(b_2^2 + b_3^2) + a_2^2(b_1^2 + b_3^2) + a_3^2(b_1^2 + b_2^2) - 2(a_1a_2b_1b_2 + a_1a_3b_1b_3 + a_2a_3b_2b_3) \\ &= |\underline{\mathbf{a}}|^2\,|\underline{\mathbf{b}}|^2 - 2(a_1a_2b_1b_2 + a_1a_3b_1b_3 + a_2a_3b_2b_3) - a_1^2b_1^2 - a_2^2b_2^2 - a_3^2b_3^2 \\ &= |\underline{\mathbf{a}}|^2\,|\underline{\mathbf{b}}|^2 - (a_1b_1 + a_2b_2 + a_3b_3)^2 \\ &= |\underline{\mathbf{a}}|^2\,|\underline{\mathbf{b}}|^2 - (\underline{\mathbf{a}} \cdot \underline{\mathbf{b}})^2 \\ &= |\underline{\mathbf{a}}|^2\,|\underline{\mathbf{b}}|^2(1 - \cos^2\theta) \\ &= |\underline{\mathbf{a}}|^2\,|\underline{\mathbf{b}}|^2|\sin\theta|^2 \\ \Rightarrow \quad |\underline{\mathbf{c}}| &= |\underline{\mathbf{a}}|\,|\underline{\mathbf{b}}|\sin\theta, \quad \text{if } \theta \text{ is minor angle between } \underline{\mathbf{a}} \text{ and } \underline{\mathbf{b}} \end{aligned}$$

Thus $\underline{\mathbf{c}}$ has the same magnitude as the geometrically defined vector product. Since $\underline{\mathbf{c}}$ is perpendicular to $\mathrm{P}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$, we also know $\underline{\mathbf{c}}$ is parallel to the geometrically defined vector product. And if $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$ are parallel then $\underline{\mathbf{c}}$ is zero. Thus if $\underline{\mathbf{c}}$ always obeys the RH rule it must be identical to the geometrically defined vector product of $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$.

If we define our algebraic vector product as the $+\underline{\mathbf{c}}$ value above, as given by (13), then by considering the coordinate vectors we obtain the following within EV, for our algebraic vector product :

$$\underline{\mathbf{u}} \times \underline{\mathbf{v}} = \underline{\mathbf{w}},\ \underline{\mathbf{v}} \times \underline{\mathbf{w}} = \underline{\mathbf{u}},\ \underline{\mathbf{w}} \times \underline{\mathbf{u}} = \underline{\mathbf{v}} \quad (14)$$

so if we use a RH CCS then these three vector products obey the RH rule, but if we use a LH CCS they obey the LH rule, since the unit basis vectors $\underline{\mathbf{u}}$, $\underline{\mathbf{v}}$, and $\underline{\mathbf{w}}$ are always directed along the positive x, y, and z axes. To make this algebraic vector product obey the RH rule with a LH CCS, at least for these three particular products in (14), we could define it as $-\underline{\mathbf{c}}$, ie. take the coordinate vector to be minus of $\begin{vmatrix} \underline{\mathbf{i}} & \underline{\mathbf{j}} & \underline{\mathbf{k}} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$. Thus we have a definition for an algebraic vector product in terms of a RH CCS and another definition in terms of LH CCS, namely $\pm\underline{\mathbf{c}}$ respectively. These definitions obey the RH rule for the vectors $\underline{\mathbf{u}}$, $\underline{\mathbf{v}}$, $\underline{\mathbf{w}}$ but we haven't shown the RH rule applies generally. For consistency the definition would have to result in the same EV vector whichever CCS we chose, ie. for all possible orientations of CCS within 3D space. However in §4.1.4 we will see this consistency must be present and the RH rule always obeyed since irrespective of the CCS used the vector produced must be equal to the EV vector product defined in §3.4.

4.1.4 EV Vector Product Expressed In Cartesian Coordinates

Unlike in the case of scalar product in §4.1.1 the formula for vector product in terms of Cartesian coordinates is different for RH CCS and LH CCS — the two formulae differ by a minus sign. Note in this section we are asking what the geometric EV vector product would look like in terms of a CCS, we are not at this point formally defining vector product on $\mathbb{R}^3$ (see §4.2).

First of all suppose the CCS is RH. The definition of the vector product in §3.4 then gives us :

$$\underline{\mathbf{u}} \times \underline{\mathbf{v}} = \underline{\mathbf{w}},\ \underline{\mathbf{v}} \times \underline{\mathbf{w}} = \underline{\mathbf{u}},\ \underline{\mathbf{w}} \times \underline{\mathbf{u}} = \underline{\mathbf{v}}$$

where $\underline{\mathbf{u}}$, $\underline{\mathbf{v}}$, $\underline{\mathbf{w}}$ are the standard basis vectors of CCS. Thus if $\underline{\mathbf{a}}, \underline{\mathbf{b}} \in$ EV have coordinate vectors $(a_1, a_2, a_3), (b_1, b_2, b_3) \in \mathbb{R}^3$ wrt $\underline{\mathbf{u}}$, $\underline{\mathbf{v}}$, $\underline{\mathbf{w}}$ then from the properties 1–5 of vector product in §3.4 we have :

$$\begin{aligned} \underline{\mathbf{a}} \times \underline{\mathbf{b}} \quad \text{in EV} \quad &= (a_1\,\underline{\mathbf{u}} + a_2\,\underline{\mathbf{v}} + a_3\,\underline{\mathbf{w}}) \times (b_1\,\underline{\mathbf{u}} + b_2\,\underline{\mathbf{v}} + b_3\,\underline{\mathbf{w}}) \\ &= (a_2b_3 - a_3b_2)\,\underline{\mathbf{u}} - (a_1b_3 - a_3b_1)\,\underline{\mathbf{v}} + (a_1b_2 - a_2b_1)\,\underline{\mathbf{w}} \\ \Rightarrow \quad \underline{\mathbf{a}} \times \underline{\mathbf{b}} \quad \text{in EV} \quad &= \begin{vmatrix} \underline{\mathbf{u}} & \underline{\mathbf{v}} & \underline{\mathbf{w}} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \qquad \text{(RH CCS)} \end{aligned} \tag{15}$$

In the case where CCS is LH we have :

$$\underline{\mathbf{u}} \times \underline{\mathbf{v}} = -\underline{\mathbf{w}},\ \underline{\mathbf{v}} \times \underline{\mathbf{w}} = -\underline{\mathbf{u}},\ \underline{\mathbf{w}} \times \underline{\mathbf{u}} = -\underline{\mathbf{v}}$$

and hence we obtain :

$$\underline{\mathbf{a}} \times \underline{\mathbf{b}} \quad \text{in EV} \quad = \quad -\begin{vmatrix} \underline{\mathbf{u}} & \underline{\mathbf{v}} & \underline{\mathbf{w}} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \qquad \text{(LH CCS)} \tag{16}$$

Taking components of (15) and (16) wrt CCS the RHS's become $\pm \begin{vmatrix} \underline{\mathbf{i}} & \underline{\mathbf{j}} & \underline{\mathbf{k}} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$, and so this coordinate vector is the same as for the algebraic vector product defined in §4.1.3, ie. $\pm\underline{\mathbf{c}}$ for RH CCS and LH CCS respectively, which proves the algebraically defined vector product of §4.1.3 and the geometrically defined vector product of §3.4 are the same, irrespective of which CCS is used for the former.

4.1.5 Scalar Triple Product Expressed In Cartesian Coordinates

If $\underline{\mathbf{a}}, \underline{\mathbf{b}}, \underline{\mathbf{c}} \in$ EV have coordinate vectors $(a_1, a_2, a_3), (b_1, b_2, b_3), (c_1, c_2, c_3) \in \mathbb{R}^3$ wrt standard basis vectors $\underline{\mathbf{u}}, \underline{\mathbf{v}}, \underline{\mathbf{w}} \in$ EV for RH CCS then applying the formulae in §4.1.1 and §4.1.4 we obtain :

$$\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) \quad \text{in EV} \quad = \quad \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} \qquad \text{(RH CCS)} \tag{17}$$

since from equation (15) the coordinate vector of $\underline{\mathbf{b}} \times \underline{\mathbf{c}}$ is $\left(\begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix}, -\begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix}, \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix} \right)$.

If instead we had a LH CCS then we would obtain :

$$\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) \quad \text{in EV} \quad = \quad -\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} \qquad \text{(LH CCS)} \tag{18}$$

4.2 Scalar And Vector Product Operations For $\mathbb{R}^3$

Scalar and vector product operations can be formally defined on $\mathbb{R}^3$ so that they are compatible with the EV scalar and vector products wrt a RH CCS — we just base them on equations (5) and (15), ie.

$$\underline{\mathbf{a}} \cdot \underline{\mathbf{b}} = a_1 b_1 + a_2 b_2 + a_3 b_3, \tag{19}$$

$$\underline{\mathbf{a}} \times \underline{\mathbf{b}} = \begin{vmatrix} \underline{\mathbf{i}} & \underline{\mathbf{j}} & \underline{\mathbf{k}} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \tag{20}$$

for $\underline{\mathbf{a}} = (a_1, a_2, a_3) \in \mathbb{R}^3$ and $\underline{\mathbf{b}} = (b_1, b_2, b_3) \in \mathbb{R}^3$, and $\underline{\mathbf{i}}$, $\underline{\mathbf{j}}$, $\underline{\mathbf{k}}$ are the unit vectors (1, 0, 0), (0, 1, 0), and (0, 0, 1) in $\mathbb{R}^3$. For example these are given as the initial definitions of scalar and vector product in the Calculus textbook by Swokowski [12].

From these definitions we then have :

$$\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) \quad \text{in } \mathbb{R}^3 \qquad = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} \tag{21}$$

It is important to note definitions (19) and (20) and equation (21) are purely algebraic and don't make any reference to geometric concepts at all or refer to any CCS, ie. they are 'a priori', even though the original motivation for making them came from geometry.

From §4.1.1, the $\mathbb{R}^3$ scalar product (19) is compatible with the EV scalar product defined in §3.2 for both RH and LH CCS in the sense that to calculate the EV scalar product we translate to the coordinate vectors in $\mathbb{R}^3$ wrt the CCS then calculate (19). Note that we have to use the standard basis vectors for the CCS (§4.1).

From §4.1.4, the $\mathbb{R}^3$ vector product (20) is compatible with the EV vector product defined in §3.4 for a RH CCS in the sense that to calculate the EV vector product we translate to the coordinate vectors in $\mathbb{R}^3$ wrt RH CCS then calculate (20), then translate back to a vector in EV. If we use a LH CCS then from (16) we would need to multiply the result by -1 to get the correct EV vector product. Note that we have to use the standard basis vectors for the CCS (§4.1).

It is easy to check that EV scalar product properties (1), (2), and (4) (commutative, multiplicative, and distributive) in §3.2 also hold for the $\mathbb{R}^3$ scalar product (19), and that properties (1), (3), (5) of the EV vector product in §3.4 (anti-commutative, multiplicative, and distributive) also hold for the $\mathbb{R}^3$ vector product (20). The differences that arise as we move from EV to $\mathbb{R}^3$, and then subsequently to $\mathbb{F}^3$ for a general field $\mathbb{F}$, are discussed in §9.2.

Combining equations (17) and (21) the $\mathbb{R}^3$ scalar triple product is compatible with the EV scalar triple product for RH CCS in the sense that :

$$\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) \quad \text{in EV} \quad = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = \underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) \quad \text{in } \mathbb{R}^3 \qquad \text{(RH CCS)} \tag{22}$$

But by (18) we must multiply by -1 if we use a LH CCS :

$$\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) \quad \text{in EV} \quad = -\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = -\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) \quad \text{in } \mathbb{R}^3 \qquad \text{(LH CCS)} \tag{23}$$

4.3 Right And Left Handed Systems In $\mathbb{R}^3$

Definition 9. For $\underline{\mathbf{a}}, \underline{\mathbf{b}}, \underline{\mathbf{c}} \in \mathbb{R}^3$ with $\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) \neq 0$ we define $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ to be a right-handed system if $\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) > 0$ and to be a left-handed system if $\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) < 0$. The handedness is undefined if $\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) = 0$.

□

With this definition it follows from (22) and Theorem 5 that $\underline{\mathbf{a}}, \underline{\mathbf{b}}, \underline{\mathbf{c}} \in \mathbb{R}^3$ are RH iff when they are plotted in a RH CCS within EV then they are RH according to Definition 3.

And it follows from (23) and Theorem 5 that $\underline{\mathbf{a}}, \underline{\mathbf{b}}, \underline{\mathbf{c}} \in \mathbb{R}^3$ are RH iff when they are plotted in a LH CCS within EV then they are LH according to Definition 3.

From (22) and (23), for both RH and LH CCS the condition $\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) = 0$ in $\mathbb{R}^3$ is equivalent to $\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) = 0$ in EV, which by Theorem 5 is in turn equivalent to $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ containing a zero vector or being coplanar (when they are plotted in EV via the CCS), ie. the same as the condition in Definition 3 for handedness in EV to be undefined.

4.4 Skew Coordinate System

Given any three non-zero non-coplanar vectors $\underline{\mathbf{a}}, \underline{\mathbf{b}}, \underline{\mathbf{c}} \in \mathrm{EV}$ a skew coordinate system CS$'$ can be constructed as shown in Figure 15, with $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ directed respectively along the x', y', z' axes, and the origin of the axis system placed at an arbitrary point O in ES. Coordinates x', y', z' of a point P wrt CS$'$ are determined as follows :

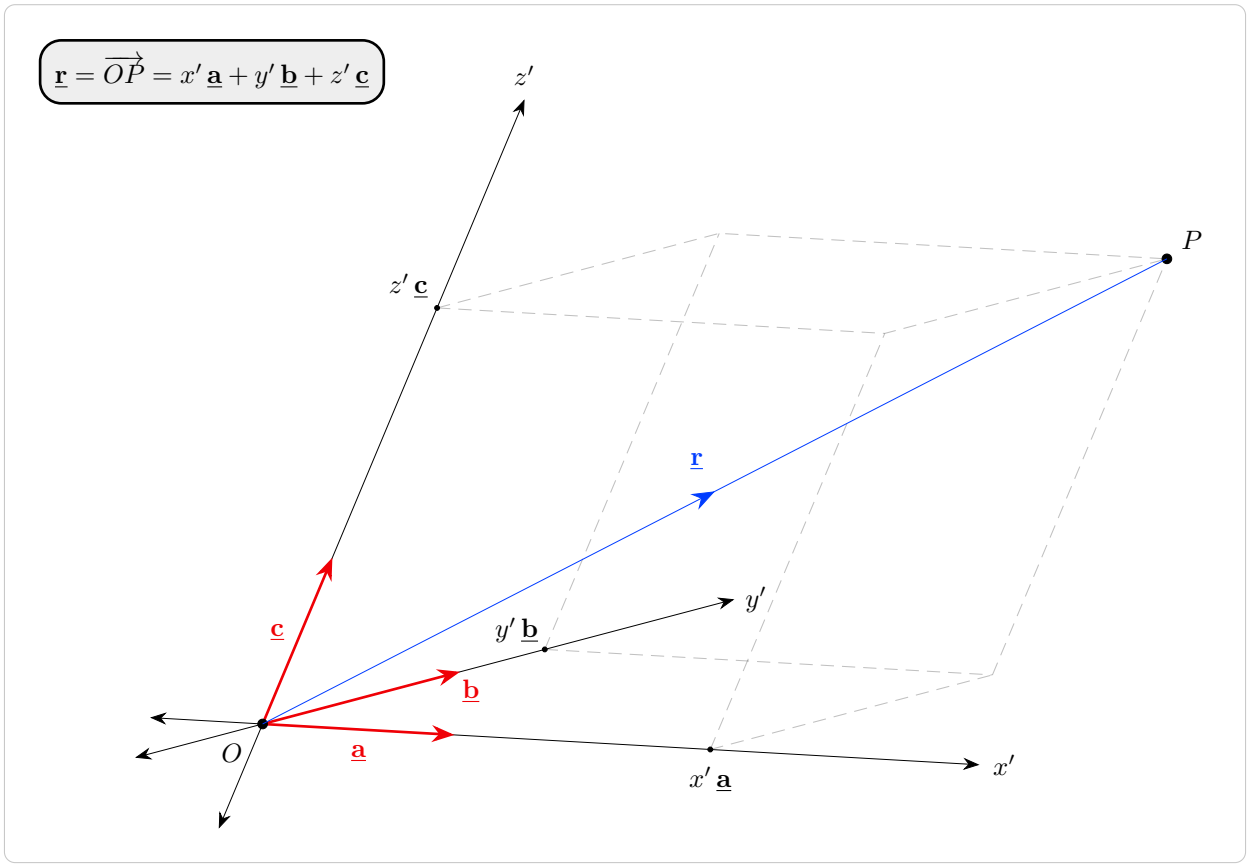
1. construct a parallelopiped with interior diagonal $\overrightarrow{OP}$ and faces parallel to the $y'z'$, $x'z'$, and $x'y'$ coordinate planes,
2. determine x', y', z' such that the vectors $x'\,\underline{\mathbf{a}}$, $y'\,\underline{\mathbf{b}}$, $z'\,\underline{\mathbf{c}}$ are respectively the position vectors of the intersection points of this parallelopiped with the x', y', z' axes.

Then as with CCS in §4.1, if $\underline{\mathbf{r}}$ is the position vector $\overrightarrow{OP}$ of P we have :

$$\underline{\mathbf{r}} = x'\,\underline{\mathbf{a}} + y'\,\underline{\mathbf{b}} + z'\,\underline{\mathbf{c}}$$

where $(x', y', z') \in \mathbb{R}^3$ is the unique coordinate vector of $\underline{\mathbf{r}}$ wrt $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$, where the vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ form a basis for the vector space EV. Since $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ are not necessarily unit vectors note x', y', z' are not in general the distances of the above intersection points from O. As with CCS, CS$'$ can be translated in any fashion and the coordinates of $\underline{\mathbf{r}}$ will remain the same (though coordinate vectors of points will change), and the mapping from vector $\underline{\mathbf{r}} \in \mathrm{EV}$ to its coordinate vector $(x', y', z') \in \mathbb{R}^3$ is a vector space isomorphism from EV to $\mathbb{R}^3$.

Figure 15: Skew Coordinate System Based On Vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$



An interesting special case of the general skew CS above is a second CCS which is obtained from the first by a general 3D rotation within ES. The positioning of such a rotated CCS with reference to the first can be described using a set of three angles called Euler angles (Appendix B).

4.5 Transformation Of Coordinates From Cartesian Coordinate System To Skew Coordinate System

To avoid overly elaborate notation below it is convenient to redefine notations used for EV vectors and $\mathbb{R}^3$ vectors at certain points, so for example we could have a vector $\underline{\mathbf{a}} \in \text{EV}$ but then redefine $\underline{\mathbf{a}}$ to mean the coordinate vector in $\mathbb{R}^3$ of $\underline{\mathbf{a}}$ wrt some given CCS.

Theorem 8. Let $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ be three arbitrary non-zero, non-coplanar vectors in EV. Let $\underline{\mathbf{r}} \in \text{EV}$ have coordinate vector $\underline{\mathbf{s}} = (x, y, z) \in \mathbb{R}^3$ wrt Cartesian coordinate system CCS (RH or LH) with its standard basis vectors, and coordinate vector $\underline{\mathbf{s}}' = (x', y', z') \in \mathbb{R}^3$ wrt skew coordinate system CS$'$ with basis vectors $\underline{\mathbf{a}}, \underline{\mathbf{b}}, \underline{\mathbf{c}} \in \text{EV}$. Then $\underline{\mathbf{s}}'$ is related to $\underline{\mathbf{s}}$ by :

$$\underline{\mathbf{s}}' \quad = \quad M\,\underline{\mathbf{s}}\,,$$

where M is the 3×3 matrix over $\mathbb{R}$ given by :

$$M \quad = \quad \frac{1}{\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}})} \begin{bmatrix} \underline{\mathbf{b}} \times \underline{\mathbf{c}} \\ \underline{\mathbf{c}} \times \underline{\mathbf{a}} \\ \underline{\mathbf{a}} \times \underline{\mathbf{b}} \end{bmatrix},$$

where $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ have been converted to their coordinate vectors in $\mathbb{R}^3$ wrt CCS.

Proof. Applying Theorem 1 of §3.2.1 to the line of the x'-axis and the $y'z'$-plane of Figure 15 we have from equation (1) of §3.2.1 :

$$x' = \frac{(\underline{\mathbf{r}} \cdot \underline{\mathbf{n}}_{y'z'})}{(\underline{\mathbf{a}} \cdot \underline{\mathbf{n}}_{y'z'})},$$

and applying it to the y'-axis and the $x'z'$-plane we have :

$$y' = \frac{(\underline{\mathbf{r}} \cdot \underline{\mathbf{n}}_{x'z'})}{(\underline{\mathbf{b}} \cdot \underline{\mathbf{n}}_{x'z'})},$$

and applying it to the z'-axis and the $x'y'$-plane we have :

$$z' = \frac{(\underline{\mathbf{r}} \cdot \underline{\mathbf{n}}_{x'y'})}{(\underline{\mathbf{c}} \cdot \underline{\mathbf{n}}_{x'y'})},$$

where $\underline{\mathbf{n}}_{y'z'}$ is a normal to the $y'z'$-plane, $\underline{\mathbf{n}}_{x'z'}$ is a normal to the $x'z'$-plane, and $\underline{\mathbf{n}}_{x'y'}$ is a normal to the $x'y'$-plane. Note the denominators above are non-zero due to the non-coplanarity of $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$. Since from §3.2.1 these do not need to be unit normals we can take :

$$\underline{\mathbf{n}}_{y'z'} = \underline{\mathbf{b}} \times \underline{\mathbf{c}}, \qquad \underline{\mathbf{n}}_{x'z'} = \underline{\mathbf{c}} \times \underline{\mathbf{a}}, \qquad \underline{\mathbf{n}}_{x'y'} = \underline{\mathbf{a}} \times \underline{\mathbf{b}}$$

giving :

$$x' = \frac{1}{\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}})}\,(\underline{\mathbf{b}} \times \underline{\mathbf{c}}) \cdot \underline{\mathbf{r}}, \qquad y' = \frac{1}{\underline{\mathbf{b}} \cdot (\underline{\mathbf{c}} \times \underline{\mathbf{a}})}\,(\underline{\mathbf{c}} \times \underline{\mathbf{a}}) \cdot \underline{\mathbf{r}}, \qquad z' = \frac{1}{\underline{\mathbf{c}} \cdot (\underline{\mathbf{a}} \times \underline{\mathbf{b}})}\,(\underline{\mathbf{a}} \times \underline{\mathbf{b}}) \cdot \underline{\mathbf{r}}.$$

But from Theorem 6 we have :

$$\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) = \underline{\mathbf{b}} \cdot (\underline{\mathbf{c}} \times \underline{\mathbf{a}}) = \underline{\mathbf{c}} \cdot (\underline{\mathbf{a}} \times \underline{\mathbf{b}})$$

thus :

$$\underline{\mathbf{s}}' \quad = \quad \frac{1}{\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}})} \begin{bmatrix} \underline{\mathbf{r}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) \\ \underline{\mathbf{r}} \cdot (\underline{\mathbf{c}} \times \underline{\mathbf{a}}) \\ \underline{\mathbf{r}} \cdot (\underline{\mathbf{a}} \times \underline{\mathbf{b}}) \end{bmatrix} \tag{24}$$

The four scalar triple products in (24) involve the EV vectors $\underline{\mathbf{r}}$, $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ and each evaluates to a real number. We can convert these EV vectors to $\mathbb{R}^3$ vectors by applying the compatibility of the scalar and vector product operations on $\mathbb{R}^3$ with the scalar and vector product operations on EV as described in §4.2. Note this requires us to be using the standard basis vectors in CCS. So in (24) convert $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ to their coordinate vectors in $\mathbb{R}^3$ wrt CCS.

Then if CCS is RH :

- the four vector product operations in (24) look exactly the same as before except they produce a result in $\mathbb{R}^3$,
- the scalar triple product in the denominator looks exactly the same as before and produces the same real number as before,
- the three scalar products with $\underline{\mathbf{r}}$ become scalar product operations with $\underline{\mathbf{r}}$ replaced by its coordinate vector $\underline{\mathbf{s}}$ wrt CCS.

If CCS is LH the four vector product operations in (24) all acquire a minus sign in front of them (from §4.2), and these minus signs thus cancel each other out, leaving the same form as in the RH case.

Thus whether the CCS we are using is RH or LH, with $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ now denoting coordinate vectors in $\mathbb{R}^3$ wrt CCS, equation (24) becomes :

$$\underline{\mathbf{s}}' = \frac{1}{\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}})} \begin{bmatrix} (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) \cdot \underline{\mathbf{s}} \\ (\underline{\mathbf{c}} \times \underline{\mathbf{a}}) \cdot \underline{\mathbf{s}} \\ (\underline{\mathbf{a}} \times \underline{\mathbf{b}}) \cdot \underline{\mathbf{s}} \end{bmatrix} = \frac{1}{\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}})} \begin{bmatrix} (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) \\ (\underline{\mathbf{c}} \times \underline{\mathbf{a}}) \\ (\underline{\mathbf{a}} \times \underline{\mathbf{b}}) \end{bmatrix} \underline{\mathbf{s}}$$

as required.

QED

4.5.1 General 3D Rotation Of CCS

When the skew CS is a general 3D rotation of the original CCS (and hence is another CCS) the matrix M is a special kind of matrix in $\mathbb{R}^{3\times3}$ called a right-handed orthogonal matrix [21], ie. $M = \begin{bmatrix} \underline{\mathbf{a}} \\ \underline{\mathbf{b}} \\ \underline{\mathbf{c}} \end{bmatrix}$ with the following properties :

- $|\underline{\mathbf{a}}| = |\underline{\mathbf{b}}| = |\underline{\mathbf{c}}| = 1$
- $\underline{\mathbf{a}} \cdot \underline{\mathbf{b}} = \underline{\mathbf{b}} \cdot \underline{\mathbf{c}} = \underline{\mathbf{c}} \cdot \underline{\mathbf{a}} = 0$
- $\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) = 1$

This is the case even when the CCS we are rotating is LH.

The first two of the above properties mean $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ are 'orthonormal' in $\mathbb{R}^3$, ie. mutually perpendicular unit vectors when plotted in EV via a CCS, and together they imply $\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) = \pm 1$. The third property means $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ form a right-handed system in $\mathbb{R}^3$ according to Definition 9.

In the RH case it is clear that $\underline{\mathbf{b}} \times \underline{\mathbf{c}} = \underline{\mathbf{a}}$, $\underline{\mathbf{c}} \times \underline{\mathbf{a}} = \underline{\mathbf{b}}$, $\underline{\mathbf{a}} \times \underline{\mathbf{b}} = \underline{\mathbf{c}}$, and $\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) = 1$. Thus we have $M = \begin{bmatrix} \underline{\mathbf{a}} \\ \underline{\mathbf{b}} \\ \underline{\mathbf{c}} \end{bmatrix}$.

To see why the above still holds for a general 3D rotation of a LH CCS, we note that in the LH case we have $\underline{\mathbf{b}} \times \underline{\mathbf{c}} = -\underline{\mathbf{a}}$ in EV, but on translating this to $\mathbb{R}^3$ we obtain $\underline{\mathbf{b}} \times \underline{\mathbf{c}} = \underline{\mathbf{a}}$, due to the left-handedness (§4.2), and similarly $\underline{\mathbf{c}} \times \underline{\mathbf{a}} = \underline{\mathbf{b}}$ and $\underline{\mathbf{a}} \times \underline{\mathbf{b}} = \underline{\mathbf{c}}$ still hold in $\mathbb{R}^3$ in the LH case. Also for LH CCS, $\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) = -1$ in EV (Theorem 5 tells us it must be negative), but by (23) this means $\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) = 1$ in $\mathbb{R}^3$.

Thus in both the RH and LH cases we have :

$$M = \frac{1}{\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}})} \begin{bmatrix} \underline{\mathbf{b}} \times \underline{\mathbf{c}} \\ \underline{\mathbf{c}} \times \underline{\mathbf{a}} \\ \underline{\mathbf{a}} \times \underline{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \underline{\mathbf{a}} \\ \underline{\mathbf{b}} \\ \underline{\mathbf{c}} \end{bmatrix}$$

Conversely all RH orthogonal matrices in $\mathbb{R}^{3\times3}$ derive from such rotations, since their row vectors represent mutually perpendicular unit vectors in EV, forming a RH system or a LH system according as we are plotting within a RH CSS or a LH CCS respectively (see comments in §4.3). In Appendix B.2 it is shown how any RH orthogonal matrix in $\mathbb{R}^{3\times3}$ can be expressed as product of three elementary rotation matrices based on Euler angles.

5 Linear Independence

In this article we make use of the notion of linear independence (LI) which applies to general vector spaces over $\mathbb{F}$ (for example as defined in [10, p108]). For three vectors in $\mathbb{R}^3$ this is the property which corresponds to the property of three vectors in EV being non-zero and non-coplanar. The formal definition is :

Definition 10. In vector space V over field $\mathbb{F}$ vectors $\underline{\mathbf{v}}_1, \underline{\mathbf{v}}_2, \ldots, \underline{\mathbf{v}}_k$ are linearly independent (LI) if $\lambda_1 \underline{\mathbf{v}}_1 + \lambda_2 \underline{\mathbf{v}}_2 + \cdots + \lambda_k \underline{\mathbf{v}}_k = \underline{\mathbf{0}} \implies \lambda_1 = \lambda_2 = \cdots = \lambda_k = 0$, ie. the only way to express the zero vector as a linear combination of $\underline{\mathbf{v}}_1, \underline{\mathbf{v}}_2, \ldots, \underline{\mathbf{v}}_k$ is with all zero coefficients. The vectors are defined to be linearly dependent (LD) if they are not linearly independent.

□

EV and $\mathbb{R}^3$ are both examples of vector spaces to which this definition applies.

The following two theorems follow easily from the definition of a vector plane in §3.1 and the vector space isomorphism between EV and $\mathbb{R}^3$ described in §4.1 and are stated without proof.

Theorem 9. In the vector space EV over $\mathbb{R}$, three vectors are LI iff they are non-zero and non-coplanar.

QED

Theorem 10. Three vectors in the vector space $\mathbb{R}^3$ are LI iff when plotted in EV using a given CCS (or indeed skew CS, §4.4) they are LI as vectors in EV.

QED

6 Matrices

In this article we consider matrices over a general field $\mathbb{F}$, as in [10, p26]. For example $\mathbb{F}$ can be $\mathbb{R}$, $\mathbb{Q}$, or $\mathbb{C}$ or it could be a finite field such as the set $\mathbb{Z}_p$ of integers mod p (aka congruence classes mod p), as described in [11, p40-45].

We commence with 3×3 matrices over $\mathbb{R}$ in §7 where we develop the matrix inverse geometrically. Then we consider 3×3 matrices over a general field $\mathbb{F}$ in §10 where we develop the matrix inverse purely algebraically (ie. in 'a priori' fashion, as per Gauss's quote in §2). For 3×3 matrices over $\mathbb{R}$, §10 thus contains an alternative proof to the geometric proof of §7, which is a purely 'a priori' algebraic proof that does not require any geometry.

We assume the following basic properties and definitions for matrices over the general field $\mathbb{F}$, without giving proof — these are all very straightforward to prove. The goal of this article is to determine the conditions under which the inverse of a square matrix in $\mathbb{F}^{n\times n}$ exists and to give the form of the inverse.

Definition 11. A left inverse of a square matrix $A \in \mathbb{F}^{n\times n}$ is a square matrix $B \in \mathbb{F}^{n\times n}$ such that $BA = I$, where I is the identity matrix in $\mathbb{F}^{n\times n}$. A right inverse of $A \in \mathbb{F}^{n\times n}$ is a square matrix $B \in \mathbb{F}^{n\times n}$ such that $AB = I$. An inverse of $A \in \mathbb{F}^{n\times n}$ is a square matrix $B \in \mathbb{F}^{n\times n}$ such that $AB = BA = I$, ie. both a left and a right inverse. A is invertible if it has an inverse, otherwise it is singular.

□

Definition 12. The transpose of a matrix $A \in \mathbb{F}^{m\times n}$ is the matrix $A^T \in \mathbb{F}^{n\times m}$ whose columns are the rows of A (and hence whose rows are the columns of A).

□

We use the term 'conformable' to mean that the dimensions of a matrix are suitable for some matrix operation, for example addition, multiplication, or inversion [22]. The following properties of matrices are assumed :

1. $\mathbb{F}^{m\times n}$ forms an abelian group under addition wrt which multiplication by any conformable matrix is distributive
2. $\mathbb{F}^{m\times n}$ has a scalar multiplication operation by elements of $\mathbb{F}$ wrt which the additive abelian group of $\mathbb{F}^{m\times n}$ becomes a vector space over $\mathbb{F}$ [10, p96]
3. matrix multiplication of conformable matrices is associative, so we may write a product of matrices without brackets, such as in $A_1A_2\cdots A_k$
4. matrix multiplication is non-commutative
5. $(A_1A_2\cdots A_k)^T = A_k^T\cdots A_2^TA_1^T$
6. the inverse of $A \in \mathbb{F}^{n\times n}$ is unique if it exists
7. if $A \in \mathbb{F}^{n\times n}$ is invertible then any left inverse equals the inverse and any right inverse equals the inverse
8. if $A_i \in \mathbb{F}^{n\times n}$ are invertible then $A_1A_2\cdots A_k$ is invertible with inverse $A_k^{-1}\cdots A_2^{-1}A_1^{-1}$
9. if A is invertible then A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$
10. if $A, B \in \mathbb{F}^{m\times n}$ satisfy $A\,\underline{\mathbf{v}} = B\,\underline{\mathbf{v}}, \forall\ \underline{\mathbf{v}} \in \mathbb{F}^n$ then $A = B$. In particular if $A \in \mathbb{F}^{n\times n}$ and $A\,\underline{\mathbf{v}} = \underline{\mathbf{v}}, \forall\ \underline{\mathbf{v}} \in \mathbb{F}^n$ then $A = I$.

Note that unlike the above properties the statement that the existence of a left inverse (or right inverse) implies the inverse exists is not trivial to prove — we require to develop some of the properties of the matrix inverse to establish this (see corollaries to Theorems 21 and 40).

For the $n \times n$ case some further matrix theory is required on echelon forms and elementary row operations and this is given §12.5.

7 Geometric Derivation Of Inverse Of 3 x 3 Matrix Over $\mathbb{R}$

In this section we derive the form of the inverse but its further properties are deferred until §10 where we cover the 3×3 matrix over a general field $\mathbb{F}$ using purely algebraic methods (thus providing an independent non-geometric proof to Theorem 11 below.).

Definition 13. Let $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ be three arbitrary vectors in $\mathbb{R}^3$ and let A be the 3×3 matrix $\left[\begin{array}{ccc} \underline{\mathbf{a}} & \underline{\mathbf{b}} & \underline{\mathbf{c}} \end{array} \right]$ over $\mathbb{R}$. Then the adjugate matrix $\operatorname{adj} A$ of A is the 3×3 matrix over $\mathbb{R}$ given by :

$$\operatorname{adj} A \quad = \quad \left[\begin{array}{c} \underline{\mathbf{b}} \times \underline{\mathbf{c}} \\ \underline{\mathbf{c}} \times \underline{\mathbf{a}} \\ \underline{\mathbf{a}} \times \underline{\mathbf{b}} \end{array} \right]$$

and the determinant $\det A$ of A is the real number given by :

$$\det A \quad = \quad \underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}})$$

□

Since this definition of $\det A$ is using the $\mathbb{R}^3$ versions of the scalar and vector product it follows from equation (21) in §4.2 that in terms of the earlier definition of determinant notation using the vertical bars in equation (7) of §4.1.2 we have :

$$\det A = \left| \begin{array}{ccc} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{array} \right| = \left| \begin{array}{c} \underline{\mathbf{a}} \\ \underline{\mathbf{b}} \\ \underline{\mathbf{c}} \end{array} \right| = \left| \begin{array}{ccc} \underline{\mathbf{a}} & \underline{\mathbf{b}} & \underline{\mathbf{c}} \end{array} \right| = |A| \,,$$

using the transpose property (8) of that earlier notation. Thus our definition of $\det A$ is equivalent to the earlier determinant notation $|A|$ defined by equation (7).

In the proof of the following theorem, we again re-use the same notation for EV vectors and $\mathbb{R}^3$ vectors to keep the notation simple, as we did in §4.5.

Theorem 11. Let $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ be three arbitrary vectors in $\mathbb{R}^3$ and let A be the matrix $\left[\begin{array}{ccc} \underline{\mathbf{a}} & \underline{\mathbf{b}} & \underline{\mathbf{c}} \end{array} \right]$. Then the following three statements are equivalent :

(a) A is invertible

(b) $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ are LI

(c) $\det A$ is non-zero

and in this case the inverse of A is given by :

$$A^{-1} \quad = \quad \frac{1}{\det A} \operatorname{adj} A$$

Proof. (a) ⇒ (b). If one of $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ was the zero vector in $\mathbb{R}^3$ then $A^{-1}A$ would have a zero column, which is impossible since $A^{-1}A = I$, so $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ are all non-zero. If they were LD one would be a linear combination of the other two, WLOG suppose that to be $\underline{\mathbf{c}}$. Then each column of AA^{-1} would be a linear combination of $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$. Then plotting these three columns in EV wrt a CCS would mean they were coplanar, which is impossible since $AA^{-1} = I = \left[\begin{array}{ccc} \underline{\mathbf{i}} & \underline{\mathbf{j}} & \underline{\mathbf{k}} \end{array} \right]$, and $\underline{\mathbf{i}}$, $\underline{\mathbf{j}}$, $\underline{\mathbf{k}}$ when plotted in a CCS can never lie in a common plane, since these become the standard basis vectors $\underline{\mathbf{u}}, \underline{\mathbf{v}}, \underline{\mathbf{w}} \in \mathrm{EV}$ for the CCS (§4.1). Thus $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ are LI.

(b) ⇒ (a). Choose a CCS. Then by Theorems 9 and 10, $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ are non-zero non-coplanar vectors when plotted in EV wrt CCS. Then by Theorem 8, and using the notation of that theorem, and allowing $\underline{\mathbf{r}}$ to range over all possible vectors in EV we have :

$$\underline{\mathbf{s}}' \quad = \quad B \, \underline{\mathbf{s}}, \; \forall \; \underline{\mathbf{s}} \in \mathbb{R}^3, \tag{25}$$

$$\text{where} \qquad B \quad = \quad \frac{1}{\det A} \operatorname{adj} A$$

where $\det A \neq 0$ due to (22) and (23) and the non-coplanarity of $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ in EV. But from definition of $\underline{\mathbf{s}}'$ and from §4.4, again allowing $\underline{\mathbf{r}}$ to range over all possible vectors in EV, we have :

$$\begin{array}{rcl} A \, \underline{\mathbf{s}}' & = & x' \, \underline{\mathbf{a}} + y' \, \underline{\mathbf{b}} + z' \, \underline{\mathbf{c}} \\ & = & \text{coordinate vector of } \underline{\mathbf{r}} \in \mathrm{EV} \text{ wrt CCS} \\ & = & \underline{\mathbf{s}} \,, \quad \text{by definition of } \underline{\mathbf{s}} \\ \Rightarrow \quad A \, \underline{\mathbf{s}}' & = & \underline{\mathbf{s}}, \; \forall \; \underline{\mathbf{s}} \in \mathbb{R}^3 \end{array} \tag{26}$$

Then combining equations (25) and (26), and noting that $\underline{\mathbf{s}}'$ as well as $\underline{\mathbf{s}}$ ranges over the whole of $\mathbb{R}^3$ as $\underline{\mathbf{r}}$ ranges over EV, we have :

$$AB\,\underline{\mathbf{s}} = \underline{\mathbf{s}},\ \forall\ \underline{\mathbf{s}} \in \mathbb{R}^3 \tag{27}$$

$$\text{and}\quad BA\,\underline{\mathbf{s}}' = \underline{\mathbf{s}}',\ \forall\ \underline{\mathbf{s}}' \in \mathbb{R}^3 \tag{28}$$

Thus both AB and BA must equal the identity matrix $I \in \mathbb{R}^{3\times 3}$, so that B is the inverse of A.

(b) $\Rightarrow$ (c). Choose a CCS. Then by Theorems 9 and 10, $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ are non-zero non-coplanar vectors when plotted in EV wrt a CCS. Then by Theorem 5 :

$$\begin{aligned} & \det A = \underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) \\ & \qquad = \pm V(\underline{\mathbf{a}}, \underline{\mathbf{b}}, \underline{\mathbf{c}}), \qquad \text{according as } \underline{\mathbf{a}},\ \underline{\mathbf{b}},\ \underline{\mathbf{c}} \text{ is RH or LH} \\ \Rightarrow\quad & \det A \neq 0 \end{aligned}$$

(c) $\Rightarrow$ (b). Again choose a CCS and plot $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ in EV wrt CCS. Equations (22) and (23) in §4.2 together with Theorem 5 then tell us that the parallelopiped is 3-dimensional, so $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ must be non-zero and non-coplanar, which from Theorems 9 and 10 implies $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ are LI both as vectors in EV and as vectors in $\mathbb{R}^3$.

Thus (a), (b), (c) are all equivalent, and when these are true the inverse B found above is the required matrix.

QED

Figuratively, by cancelling the cross terms in their numerators and denominators, we can think of the rows $\dfrac{(\underline{\mathbf{b}} \times \underline{\mathbf{c}})}{\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}})}$, $\dfrac{(\underline{\mathbf{c}} \times \underline{\mathbf{a}})}{\underline{\mathbf{b}} \cdot (\underline{\mathbf{c}} \times \underline{\mathbf{a}})}$, $\dfrac{(\underline{\mathbf{a}} \times \underline{\mathbf{b}})}{\underline{\mathbf{c}} \cdot (\underline{\mathbf{a}} \times \underline{\mathbf{b}})}$ of A^{-1} as being a kind of reciprocal operation $\dfrac{1}{\underline{\mathbf{a}}}$, $\dfrac{1}{\underline{\mathbf{b}}}$, $\dfrac{1}{\underline{\mathbf{c}}}$ on $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ respectively, so:

$$A^{-1}A = \begin{bmatrix} 1/\underline{\mathbf{a}} \\ 1/\underline{\mathbf{b}} \\ 1/\underline{\mathbf{c}} \end{bmatrix} \begin{bmatrix} \underline{\mathbf{a}} & \underline{\mathbf{b}} & \underline{\mathbf{c}} \end{bmatrix} = I.$$

According to this definition $1/\underline{\mathbf{a}}$ would be perpendicular to $\underline{\mathbf{b}}$ and $\underline{\mathbf{c}}$, $1/\underline{\mathbf{b}}$ would be perpendicular to $\underline{\mathbf{c}}$ and $\underline{\mathbf{a}}$, and $1/\underline{\mathbf{c}}$ would be perpendicular to $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$, so giving the zeros off the diagonal. Multiplying the vectors by their reciprocals would give the 1's down the diagonal.

8 Comments On The General Field $\mathbb{F}$

The arithmetic operations of addition, subtraction, multiplication, and division in a general field $\mathbb{F}$ function very similarly to those of $\mathbb{R}$ making the field a very familiar environment to work with, however there are certain differences that can arise which we need to be aware of, for example :

- there may be no order relation in the field, and so no notion of positive and negative elements in the field
- the field may be finite
- the field may have characteristic 2, which means division by 2 is not possible

The theory of the matrix inverse that we develop below applies to matrices over a general field $\mathbb{F}$. In the Linear Algebra textbook [10] the chapters 2, 3, and 4 on matrices all apply to a general field $\mathbb{F}$, and vector spaces over a general field $\mathbb{F}$ of scalars are considered, for example the vector space $\mathbb{F}^n$ over $\mathbb{F}$. The field $\mathbb{F}$ could be $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{Q}$, which are all examples of infinite fields. However $\mathbb{F}$ may also be a finite field (aka Galois field [23]), such as the field $\mathbb{Z}_p$ of integers mod p where p is a prime number, which is a field of order p. Generally finite fields always have an order which is a prime power, and for every prime power p^k there is a field of order p^k, and all fields of order p^k are isomorphic to one another. Finite fields and matrices over finite fields have many applications including in cryptography and coding theory as well as in mathematics itself, thus it is desirable to develop a theory of matrix inverses which applies to them as well as to the more familiar fields such as a $\mathbb{R}$ or $\mathbb{C}$.

In the field $\mathbb{R}$ we have the familiar order relation $\leq$ which is a 'total order' relation [14]. A general field $\mathbb{F}$ may also have an order relation and the combination of the field and its order relation is termed an 'ordered field' [24], if the order is a total order and is compatible with the field operations of multiplication and addition, ie. if $\forall\ a, b, c \in \mathbb{F}$:

- $a \leq b \Rightarrow a + c \leq b + c$, and
- $0 \leq a$ and $0 \leq b \Rightarrow 0 \leq ab$

In a general ordered field we then define the relation $\geq$ and the strict order relations, $<, >$ in terms of $\leq$ in the natural way, ie.

- $a \geq b \iff b \leq a$
- $a < b \iff a \leq b$ and $a \neq b$
- $a > b \iff b < a$

Then many properties involving the order relations can be deduced which mirror the familiar properties of $\mathbb{R}$, eg.

- the law of trichotomy [25], ie. $\forall\ a, b \in \mathbb{F}$, exactly one of (i) $a < b$, (ii) $a = b$, or (iii) $a > b$ holds
- $\forall a, b \in \mathbb{F}$, if $0 < a < b$ then $1/b < 1/a$
- $\forall\ a \in \mathbb{F}, a^2 \geq 0$
- in $\mathbb{F}$, $0 < 1$

The last of these properties $0 < 1$ seems too obvious to state but in a general field $\mathbb{F}$ it does require to be proved from the ordered field axioms. A consequence of it is that in any ordered field, we must have :

$$\ldots < -1 - 1 - 1 < -1 - 1 < -1 < 0 < 1 < 1 + 1 < 1 + 1 + 1 < \ldots \tag{29}$$

so that any ordered field $\mathbb{F}$ must be infinite.

An ordered field $\mathbb{F}$ is deemed to be 'complete' if every non-empty subset of $\mathbb{F}$ which is bounded above has a least upper bound. The field $\mathbb{R}$ has this property (Appendix C) and it is an important property in mathematical analysis. It can be shown that all complete ordered fields are isomorphic so that there is essentially only one field $\mathbb{R}$. The field $\mathbb{Q}$ lacks the completeness property owing to the irrationality of $\sqrt{2}$, the bounded set $\{q \in \mathbb{Q}^+ : q^2 < 2\}$ having an upper bound but no least upper bound in $\mathbb{Q}$.

As we observed from equation (29) it is not possible for any finite field to become an ordered field according to the above definition — an order relation may be defined on it but it can never be compatible with the field operations in the above sense. For example in $\mathbb{Z}_p$ with $p = 11$ we can see the incompatibility with addition by adding 5 to the inequality $4 \leq 7$ to give the incorrect $9 \leq 1$, and this would be the case for any total order relation $\leq$ on $\mathbb{Z}_{11}$ satisfying $0 < 1 < 2 < \ldots < 10$ (as any compatible order relation would have to do). In the case of $\mathbb{C}$ a lexicographic order relation similar to dictionary ordering could be defined by comparing the real part first and the imaginary part second to give a total order relation but this would not be compatible with the field operations on $\mathbb{C}$. No compatible order relation can exist on $\mathbb{C}$ since by equation (29) it would have to satisfy $-1 < 0$, implying $i^2 < 0$ — but we must have $a^2 \geq 0, \forall\ a \in \mathbb{F}$ for any ordered field $\mathbb{F}$.

Inner product spaces ([26], [10, p207]) are not defined over general fields but only over $\mathbb{R}$ and $\mathbb{C}$. An inner product operation $\langle \cdot, \cdot \rangle$ on a vector space V over field $\mathbb{F}$ ($= \mathbb{R}$ or $\mathbb{C}$) is a mapping from $V \times V \to \mathbb{F}$ and requires the property of 'positive definiteness' (ie. $\langle a, a \rangle > 0,\ \forall\ a \in V \setminus \{\underline{\mathbf{0}}\}$) which requires $\mathbb{F}$ to have an order relation, but this works for $\mathbb{C}$ also even though $\mathbb{C}$ has no compatible order relation since $\langle a, a \rangle$ is always real for both $\mathbb{R}$ and $\mathbb{C}$.

When we are working with matrices over a general field $\mathbb{F}$ where there may be no order relation we cannot consider whether a matrix determinant (defined for $\mathbb{R}^{3\times 3}$ using the geometric approach in Definition 13 in §7) is positive or negative since there is no notion of positive or negative without an order relation. On $\mathbb{R}^{3\times 3}$ the sign of the determinant is connected with the notion of right and left handedness of a system of three vectors (§4.3), so in $\mathbb{F}^3$ for a field $\mathbb{F}$ with no order relation the concept of RH/LH would not be well defined.

A field of characteristic 2 (denoted char $\mathbb{F} = 2$) is a field in which $1 + 1 = 0$, ie. $2 = 0$. This is the smallest possible field and consists of just two elements, 0 and 1. Generally the characteristic of a field is the least number $n \in \mathbb{N}$ such that:

$$\underbrace{1 + 1 + \cdots + 1}_{n} = 0\,,$$

or 0 if no such n exists. From equation (29) any ordered field must have char $\mathbb{F} = 0$. If the characteristic is non-zero then it must be a prime. Division by 2 is not possible for a field with char $\mathbb{F} = 2$, which can present a problem for certain operations involving matrices or linear algebra generally, eg. the chapter on quadratic forms in [10, p226] excludes fields with char $\mathbb{F} = 2$. For Theorem 29 in this article which states that the determinant is zero if there are two equal rows or two equal columns the proof is immediate when char $\mathbb{F} \neq 2$, where we can divide by 2, but a longer proof is needed to cover the char $\mathbb{F} = 2$ case.

9 Vector Space $\mathbb{F}^3$ Over General Field $\mathbb{F}$

9.1 Definition Of Scalar And Vector Product On $\mathbb{F}^3$

To facilitate the development of the matrix inverse on $\mathbb{F}^{3\times3}$, we extend the definitions in §4.2 of the scalar product operation on the vector space $\mathbb{R}^3$ to the vector space $\mathbb{F}^3$, for a general field $\mathbb{F}$. This is not intended to be an inner product on $\mathbb{F}^3$ since inner products are not well-defined for vector spaces over general fields but only over $\mathbb{R}$ and $\mathbb{C}$ ([26]), as we mentioned in §8 above.

This definition provides a notational convenience in $\mathbb{F}^3$ and helps to import the concepts we learned about matrix inverses in geometry over to the purely algebraic case of $\mathbb{F}^{3\times3}$ where there is no geometric interpretation. For the same purpose we also extend to $\mathbb{F}^3$ the definition of the vector product operation on $\mathbb{R}^3$ from §4.2.

Thus on $\mathbb{F}^3$ we define :

$$\underline{\mathbf{a}} \cdot \underline{\mathbf{b}} = a_1b_1 + a_2b_2 + a_3b_3, \tag{30}$$

$$\underline{\mathbf{a}} \times \underline{\mathbf{b}} = \begin{vmatrix} \underline{\mathbf{i}} & \underline{\mathbf{j}} & \underline{\mathbf{k}} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \tag{31}$$

where $\underline{\mathbf{a}} = (a_1, a_2, a_3)$ and $\underline{\mathbf{b}} = (b_1, b_2, b_3) \in \mathbb{F}^3$, and $\underline{\mathbf{i}}$, $\underline{\mathbf{j}}$, $\underline{\mathbf{k}}$ are the unit vectors (1, 0, 0), (0, 1, 0), and (0, 0, 1) in $\mathbb{F}^3$.

The scalar triple product on $\mathbb{F}^3$ is then given by :

$$\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} \tag{32}$$

Although these have no geometric interpretation for a general field $\mathbb{F}$ and are just formal definitions on $\mathbb{F}^3$ they are originally inspired from Euclidean geometry, specifically the representations of the EV scalar and vector products wrt a CCS in equations (5) and (15).

9.2 Comparison Of $\mathbb{F}^3$ With Vector Spaces EV And $\mathbb{R}^3$

Working in the vector space $\mathbb{F}^3$ over $\mathbb{F}$ (which includes the vector space $\mathbb{R}^3$ over $\mathbb{R}$ as a special case) is purely 'a priori' as in Gauss's quote in §2, since the 'physical model' of Euclidean geometry is not involved, and everything is deduced purely algebraically from first principles as in §9.3 and §10 below.

Within $\mathbb{R}^3$ geometric notions like direction, angle, perpendicularity, right-handedness, area, and volume just become definitions within a formal system. For example there is no Pythagoras Theorem within $\mathbb{R}^3$, only a definition of a distance function (a 'metric') between points within $\mathbb{R}^3$ which is inspired by the Pythagorean identity. Perpendicularity of vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$ within the geometric vector space EV becomes the algebraic notion of 'orthogonality' of $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$ in $\mathbb{R}^3$, defined to mean $\underline{\mathbf{a}} \cdot \underline{\mathbf{b}} = 0$. We can define the notion of RH/LH of a system of three vectors in $\mathbb{R}^3$ in terms of the scalar triple product $\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}})$ being positive/negative (§4.3). The notion of two vectors in EV being non-zero and non-parallel becomes the notion of $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$ being LI in $\mathbb{R}^3$ (LI is defined for a general vector space in Definition 10 above), and the notion of three vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ in EV being non-zero and non-coplanar becomes the notion of $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ being LI in $\mathbb{R}^3$.

Some of the concepts that carry over from EV to $\mathbb{R}^3$ do not carry over to $\mathbb{F}^3$ for the general field. The idea of orthogonality of two vectors [27] is defined for a general inner product space so it applies to $\mathbb{R}^3$ which forms an inner product space with its scalar product operation, but it does not apply to $\mathbb{F}^3$ for a general field $\mathbb{F}$ since inner product spaces are only well-defined over $\mathbb{R}$ and $\mathbb{C}$ [26]. Positive and negative scalar values are not defined for a field with no order relation such as $\mathbb{C}$ or a finite field. We have no notion of length of a vector in $\mathbb{F}^3$ over a general field — to attempt to analogize with $\mathbb{R}$ we have $\underline{\mathbf{a}} \cdot \underline{\mathbf{a}}$ as length squared but this is not even a real number, it is an element of the general field $\mathbb{F}$. The notion of RH/LH of a system of vectors in $\mathbb{F}^3$ for a general field cannot be defined as we did for $\mathbb{R}^3$ since we cannot specify the sign of $\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}})$ in the absence of an order relation.

The idea of LI (Definition 10) applies to any vector space over any field so it applies to $\mathbb{F}^n$ for any field $\mathbb{F}$ and any $n \in \mathbb{N}$, and we could if needed easily extend the scalar product to $\mathbb{F}^n$ but not the vector product which does not apply to dimensions higher than three. The situation is summarized in Table 1 below.

Table 1: Corresponding Concepts In The Vector Spaces EV, $\mathbb{R}^3$, And $\mathbb{F}^3$

EV	$\mathbb{R}^3$	$\mathbb{F}^3$
angle as a primitive concept	angle between vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$ defined analytically as $\arccos(\underline{\mathbf{a}} \cdot \underline{\mathbf{b}}/\lvert\underline{\mathbf{a}}\rvert\,\lvert\underline{\mathbf{b}}\rvert)$	not defined
vector length $\lvert\underline{\mathbf{a}}\rvert$ given by Pythagoras formula	vector length formally defined by $\lvert\underline{\mathbf{a}}\rvert = \sqrt{a_1^2 + a_2^2 + a_3^2} = \sqrt{\underline{\mathbf{a}} \cdot \underline{\mathbf{a}}}$	not defined
positive and negative scalars	as EV, defined by order relation on $\mathbb{R}$	not defined
non-zero and non-parallel vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$	LI vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$	as $\mathbb{R}^3$
non-zero and non-coplanar vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$	LI vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$	as $\mathbb{R}^3$
perpendicular vectors $\underline{\mathbf{a}}, \underline{\mathbf{b}}$, defined by right angle	orthogonal vectors, defined by $\underline{\mathbf{a}} \cdot \underline{\mathbf{b}} = 0$	not defined
RH and LH systems of vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ defined by angle sense	RH/LH defined by $\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) > 0$ or < 0	not defined
scalar product $\underline{\mathbf{a}} \cdot \underline{\mathbf{b}} = \lvert\underline{\mathbf{a}}\rvert\,\lvert\underline{\mathbf{b}}\rvert \cos\theta$	scalar product $\underline{\mathbf{a}} \cdot \underline{\mathbf{b}} = a_1b_1 + a_2b_2 + a_3b_3$	as $\mathbb{R}^3$
vector product $\underline{\mathbf{a}} \times \underline{\mathbf{b}}$: length $= \lvert\underline{\mathbf{a}}\rvert\,\lvert\underline{\mathbf{b}}\rvert \sin\theta$, direction from RH rule	vector product $\underline{\mathbf{a}} \times \underline{\mathbf{b}} = (a_2b_3 - a_3b_2, a_3c_1 - a_1c_3, a_1b_2 - a_2b_1)$	as $\mathbb{R}^3$

9.3 Properties Of Scalar And Vector Product On $\mathbb{F}^3$

It is easy to check that EV scalar product properties (1), (2), and (4) (commutative, multiplicative, and distributive) in §3.2 hold for definition (30). In the case of $\mathbb{R}^3$ (but not $\mathbb{F}^3$ generally) property (3) in §3.2 is replaced by the definition of the algebraic notion of orthogonality : vectors $\underline{\mathbf{a}}, \underline{\mathbf{b}} \in \mathbb{R}^3$ are orthogonal (denoted $\underline{\mathbf{a}} \perp \underline{\mathbf{b}}$) if $\underline{\mathbf{a}} \cdot \underline{\mathbf{b}} = 0$. The idea of orthogonality is not well-defined on $\mathbb{F}^3$ for a general field $\mathbb{F}$ (§9.2).

Properties (1), (3), (5) of the EV vector product in §3.4 (anti-commutative, multiplicative, and distributive) are straightforward to check for definition (31). However the geometric idea of 'parallel' in EV becomes the notion of 'linear dependence' (LD) in the vector space $\mathbb{F}^3$ (see Definition 10 above), so that property (4) becomes : $\underline{\mathbf{a}} \times \underline{\mathbf{b}} = \underline{\mathbf{0}} \iff \underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ are LD, which we show below. Two vectors are LD in a general vector space iff one is a scalar multiple of the other. It further follows easily from (30) and (31) that $\forall\ \underline{\mathbf{a}}, \underline{\mathbf{b}} \in \mathbb{F}^3$, $(\underline{\mathbf{a}} \times \underline{\mathbf{b}}) \cdot \underline{\mathbf{a}} = 0$, $(\underline{\mathbf{a}} \times \underline{\mathbf{b}}) \cdot \underline{\mathbf{b}} = 0$, and $\underline{\mathbf{a}} \times \underline{\mathbf{a}} = \underline{\mathbf{0}}$.

The theorems 12–17 below are required for §10 on matrix inverses in $\mathbb{F}^{3\times3}$. Numerous other properties can be derived but these are all we need for the present purpose. These purely algebraic proofs for $\mathbb{F}^3$ can be contrasted with their geometrical counterparts for EV in §3.

Theorem 12. If $\underline{\mathbf{a}}, \underline{\mathbf{b}} \in \mathbb{F}^3$, then $\underline{\mathbf{a}} \times \underline{\mathbf{b}} = \underline{\mathbf{0}}$ iff $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$ are LD.

Proof. Let $\underline{\mathbf{a}} = (a_1, a_2, a_3)$ and $\underline{\mathbf{b}} = (b_1, b_2, b_3)$.

$\Leftarrow$. WLOG suppose $\underline{\mathbf{b}}$ a scalar multiple of $\underline{\mathbf{a}}$, say $\underline{\mathbf{b}} = \lambda\,\underline{\mathbf{a}}$. Then $\underline{\mathbf{a}} \times \underline{\mathbf{b}} = \lambda\,(\underline{\mathbf{a}} \times \underline{\mathbf{a}}) = \underline{\mathbf{0}}$.

$\Rightarrow$. Let $\underline{\mathbf{a}} \times \underline{\mathbf{b}} = \underline{\mathbf{0}}$, so :

$$a_2b_3 = a_3b_2, \quad a_3b_1 = a_1b_3, \quad a_1b_2 = a_2b_1$$

The case of $\underline{\mathbf{a}} = \underline{\mathbf{0}}$ is clear so assume $\underline{\mathbf{a}} \neq \underline{\mathbf{0}}$. Then either $a_1 \neq 0$, $a_2 \neq 0$, or $a_3 \neq 0$. If $a_1 \neq 0$, we can write :

$$b_1 = \frac{b_1}{a_1}\,a_1, \quad b_2 = \frac{b_1}{a_1}\,a_2, \quad b_3 = \frac{b_1}{a_1}\,a_3$$

and so $\underline{\mathbf{b}} = \lambda\,\underline{\mathbf{a}}$ where $\lambda = b_1/a_1$, from which it follows that $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$ are LD.

The cases of $a_2 \neq 0$ and $a_3 \neq 0$ follow similarly.

QED

Theorem 13. In $\mathbb{F}^3$, $\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) = 0$ iff $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ are LD.

Proof. $\Leftarrow$. Suppose WLOG that we had $\lambda\,\underline{\mathbf{a}} + \mu\,\underline{\mathbf{b}} + \nu\,\underline{\mathbf{c}} = \underline{\mathbf{0}}$ where $\lambda \neq 0$. Then :

$$\begin{aligned} \underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) &= \left(-\frac{\mu}{\lambda}\,\underline{\mathbf{b}} - \frac{\nu}{\lambda}\,\underline{\mathbf{c}}\right) \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) \\ &= 0, \end{aligned}$$

using the distributive law.

$\Rightarrow$. Firstly consider the case where $\underline{\mathbf{b}} \times \underline{\mathbf{c}} = \underline{\mathbf{0}}$. Then by Theorem 12 above $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ are LD and thus $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ are LD.

Otherwise $\underline{\mathbf{b}} \times \underline{\mathbf{c}} = (b_2c_3 - b_3c_2, b_3c_1 - b_1c_3, b_1c_2 - b_2c_1)$ has at least one non-zero component. To show $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ are LD consider finding λ and μ such that $\underline{\mathbf{a}} = \lambda\,\underline{\mathbf{b}} + \mu\,\underline{\mathbf{c}}$. This requires :

$$\begin{aligned} a_1 &= \lambda\,b_1 + \mu\,c_1 \\ a_2 &= \lambda\,b_2 + \mu\,c_2 \\ a_3 &= \lambda\,b_3 + \mu\,c_3 \end{aligned}$$

the first two of which require :

$$\begin{bmatrix} b_1 & c_1 \\ b_2 & c_2 \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

which has the following solution if the 2×2 determinant $b_1c_2 - b_2c_1 \neq 0$:

$$\begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \frac{1}{b_1c_2 - b_2c_1} \begin{bmatrix} c_2 & -c_1 \\ -b_2 & b_1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}.$$

Thus if $b_1c_2 - b_2c_1 \neq 0$ then if the third equation $a_3 = \lambda\,b_3 + \mu\,c_3$ holds, ie. if

$$(b_1c_2 - b_2c_1)a_3 = (c_2a_1 - c_1a_2)b_3 + (-a_1b_2 + a_2b_1)c_3$$

then we have found the required λ and μ, so that $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ are LD. But from the hypothesis we have :

$$\begin{aligned} a_1(b_2c_3 - b_3c_2) + a_2(b_3c_1 - b_1c_3) + a_3(b_1c_2 - b_2c_1) &= 0 \\ \Rightarrow \qquad (b_1c_2 - b_2c_1)a_3 &= -a_1(b_2c_3 - b_3c_2) - a_2(b_3c_1 - b_1c_3) \\ &= (c_2a_1 - c_1a_2)b_3 + (-a_1b_2 + a_2b_1)c_3 \end{aligned}$$

as required. Similarly in the cases $b_3c_1 - b_1c_3 \neq 0$ and $b_2c_3 - b_3c_2 \neq 0$ we are able to find λ and μ such that $\underline{\mathbf{a}} = \lambda\,\underline{\mathbf{b}} + \mu\,\underline{\mathbf{c}}$. Hence $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ are LD.

QED

Theorem 14. In $\mathbb{F}^3$, $\underline{\mathbf{a}}, \underline{\mathbf{b}}, \underline{\mathbf{c}} \in \mathrm{sp}(\underline{\mathbf{d}}, \underline{\mathbf{e}}) \Rightarrow \underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) = 0$.

Proof. We have

$$\underline{\mathbf{a}} = \lambda_1\,\underline{\mathbf{d}} + \mu_1\,\underline{\mathbf{e}}, \quad \underline{\mathbf{b}} = \lambda_2\,\underline{\mathbf{d}} + \mu_2\,\underline{\mathbf{e}}, \quad \underline{\mathbf{c}} = \lambda_3\,\underline{\mathbf{d}} + \mu_3\,\underline{\mathbf{e}}$$

Thus

$$\begin{aligned} \underline{\mathbf{b}} \times \underline{\mathbf{c}} &= \lambda_2\mu_3\,(\underline{\mathbf{d}} \times \underline{\mathbf{e}}) + \mu_2\lambda_3\,(\underline{\mathbf{e}} \times \underline{\mathbf{d}}) \\ &= \lambda_2\mu_3\,(\underline{\mathbf{d}} \times \underline{\mathbf{e}}) - \mu_2\lambda_3\,(\underline{\mathbf{d}} \times \underline{\mathbf{e}}) \end{aligned}$$

and so

$$\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) = 0$$

since $\underline{\mathbf{a}}$ is a linear combination of $\underline{\mathbf{d}}$ and $\underline{\mathbf{e}}$.

QED

Theorem 15. In $\mathbb{F}^3$ the value of the scalar triple product $\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}})$ is unchanged after any cyclic reordering of the three vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$, ie.

$$\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) = \underline{\mathbf{c}} \cdot (\underline{\mathbf{a}} \times \underline{\mathbf{b}}) = \underline{\mathbf{b}} \cdot (\underline{\mathbf{c}} \times \underline{\mathbf{a}}).$$

Interchanging any two of the vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ in $\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}})$ flips the sign of the scalar triple product, ie.

$$\underline{\mathbf{b}} \cdot (\underline{\mathbf{a}} \times \underline{\mathbf{c}}) = \underline{\mathbf{a}} \cdot (\underline{\mathbf{c}} \times \underline{\mathbf{b}}) = \underline{\mathbf{c}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{a}}) = -\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}).$$

Proof. From (32) above we have :

$$\begin{aligned}
\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) &= a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix} \\
\underline{\mathbf{c}} \cdot (\underline{\mathbf{a}} \times \underline{\mathbf{b}}) &= c_1 \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} - c_2 \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} + c_3 \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \\
\underline{\mathbf{b}} \cdot (\underline{\mathbf{c}} \times \underline{\mathbf{a}}) &= b_1 \begin{vmatrix} c_2 & c_3 \\ a_2 & a_3 \end{vmatrix} - b_2 \begin{vmatrix} c_1 & c_3 \\ a_1 & a_3 \end{vmatrix} + b_3 \begin{vmatrix} c_1 & c_2 \\ a_1 & a_2 \end{vmatrix}
\end{aligned}$$

which by inspection are all equal (eg. gather c_1 terms in the first and compare with the second, and so on). The second part of the theorem follows from the anti-commutative property of the vector product.

QED

Theorem 16. For any $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ in $\mathbb{F}^3$, the vector triple product $\underline{\mathbf{a}} \times (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) = (\underline{\mathbf{a}} \cdot \underline{\mathbf{c}})\,\underline{\mathbf{b}} - (\underline{\mathbf{a}} \cdot \underline{\mathbf{b}})\,\underline{\mathbf{c}}$.

Proof. From (31) above :

$$\begin{aligned}
\underline{\mathbf{a}} \times (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) &= \begin{vmatrix} \underline{\mathbf{i}} & \underline{\mathbf{j}} & \underline{\mathbf{k}} \\ a_1 & a_2 & a_3 \\ b_2c_3 - b_3c_2 & b_3c_1 - b_1c_3 & b_1c_2 - b_2c_1 \end{vmatrix} \\
\Rightarrow \quad \text{component 1} &= a_2(b_1c_2 - b_2c_1) - a_3(b_3c_1 - b_1c_3) \\
&= (a_2c_2 + a_3c_3)b_1 - (a_2b_2 + a_3b_3)c_1 \\
&= (a_1c_1 + a_2c_2 + a_3c_3)b_1 - (a_1b_1 + a_2b_2 + a_3b_3)c_1 \\
&= (\underline{\mathbf{a}} \cdot \underline{\mathbf{c}})\,b_1 - (\underline{\mathbf{a}} \cdot \underline{\mathbf{b}})\,c_1 \\
&= \text{component 1 of } (\underline{\mathbf{a}} \cdot \underline{\mathbf{c}})\,\underline{\mathbf{b}} - (\underline{\mathbf{a}} \cdot \underline{\mathbf{b}})\,\underline{\mathbf{c}}
\end{aligned}$$

and similarly for components 2 and 3, thus

QED

Theorem 17. For any vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$, $\underline{\mathbf{d}}$ in $\mathbb{F}^3$, $(\underline{\mathbf{a}} \times \underline{\mathbf{b}}) \cdot (\underline{\mathbf{c}} \times \underline{\mathbf{d}}) = (\underline{\mathbf{a}} \cdot \underline{\mathbf{c}})\,(\underline{\mathbf{b}} \cdot \underline{\mathbf{d}}) - (\underline{\mathbf{b}} \cdot \underline{\mathbf{c}})\,(\underline{\mathbf{a}} \cdot \underline{\mathbf{d}})$.

Proof. We have :

$$\begin{aligned}
(\underline{\mathbf{a}} \times \underline{\mathbf{b}}) \cdot (\underline{\mathbf{c}} \times \underline{\mathbf{d}}) &= \underline{\mathbf{d}} \cdot ((\underline{\mathbf{a}} \times \underline{\mathbf{b}}) \times \underline{\mathbf{c}}), \qquad \text{by Theorem 15} \\
&= -\underline{\mathbf{d}} \cdot (\underline{\mathbf{c}} \times (\underline{\mathbf{a}} \times \underline{\mathbf{b}})) \\
&= -\underline{\mathbf{d}} \cdot ((\underline{\mathbf{c}} \cdot \underline{\mathbf{b}})\,\underline{\mathbf{a}} - (\underline{\mathbf{c}} \cdot \underline{\mathbf{a}})\,\underline{\mathbf{b}}), \qquad \text{by Theorem 16} \\
&= (\underline{\mathbf{a}} \cdot \underline{\mathbf{c}})\,(\underline{\mathbf{b}} \cdot \underline{\mathbf{d}}) - (\underline{\mathbf{b}} \cdot \underline{\mathbf{c}})\,(\underline{\mathbf{a}} \cdot \underline{\mathbf{d}})
\end{aligned}$$

QED

10 Inverse Of 3 x 3 Matrix Over General Field $\mathbb{F}$

This section makes use of the results obtained in §9.3 above and as in that section the proofs are all purely algebraic and 'a priori' and do not depend on any of the geometric sections above. However Definition 14 made below, which extends Definition 13 in §7 for 3×3 matrices over $\mathbb{R}$ to 3×3 matrices over a general field $\mathbb{F}$, is clearly inspired by the form of the inverse as determined in the geometric approach, ie. we have not done any algebraic work to arrive at Definition 14. A second way we use the geometrically determined form of the matrix inverse is in suggesting Theorem 18.

To find the form of the matrix inverse purely algebraically we would need to solve 3 systems of 3 equations in 3 unknowns. However each system would take the same form and we would just need to interchange the vectors $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ to arrive at the solution to one system from the other. We would find the 3 systems were all solvable if $\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) \neq 0$. Whether we do that algebraic work or not the properties are still all formally proved below, ie. although we 'cheat' by using the geometry to guide us we still provide full formal proof of everything claimed using algebra only, so that the proofs fully cover the case of the general field $\mathbb{F}$.

The geometric approach is more natural and intuitive whereas the algebraic approach is more 'brute force'. However the geometric method of matrix inversion came about accidentally since it was solved by a completely different question of how to transform coordinate systems — which just turned out to also solve the matrix inverse problem and hence give

the general solution to the 3×3 system of linear equations. Just as a vector in EV and a vector in $\mathbb{R}^3$ are objects within different categories so the geometric proof in §7 above and the algebraic proofs in this section below fall into different categories. The category of geometry we used above depends on physical concepts as per Gauss's quote in §2 whereas the category of algebra does not contain physical concepts and is purely 'a priori'.

The advantage of the algebraic approach is that it applies to matrices over a general field $\mathbb{F}$ and it also leads to a solution for the general $n \times n$ matrix inverse, neither of which the geometric approach can accomplish. Note also that the algebraic proof retrospectively covers the original $\mathbb{R}^{3\times3}$ case in an 'a priori' fashion that stands independently of the geometric proof of §7.

The definition below uses the $\mathbb{F}^3$ versions of the scalar and vector product operations as defined in §9.1 above.

Definition 14. Let $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ be three arbitrary vectors in $\mathbb{F}^3$ and let A be the 3×3 matrix $\left[\begin{array}{ccc} \underline{\mathbf{a}} & \underline{\mathbf{b}} & \underline{\mathbf{c}} \end{array}\right]$ over $\mathbb{F}$. Then the adjugate matrix $\operatorname{adj} A$ of A is the 3×3 matrix over $\mathbb{F}$ given by :

$$\operatorname{adj} A \quad = \quad \left[\begin{array}{c} \underline{\mathbf{b}} \times \underline{\mathbf{c}} \\ \underline{\mathbf{c}} \times \underline{\mathbf{a}} \\ \underline{\mathbf{a}} \times \underline{\mathbf{b}} \end{array}\right]$$

and the determinant $\det A$ of A is the real number given by :

$$\det A \quad = \quad \underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}})$$

□

Since this definition of $\det A$ is using the $\mathbb{F}^3$ versions of the scalar and vector product, it follows from equation (32) in §9.1, which uses the earlier definition of determinant notation using the vertical bars in equation (7), and from the transpose property of that determinant notation that :

$$\det A = \underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) = \left|\begin{array}{ccc} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{array}\right| = \left|\begin{array}{c} \underline{\mathbf{a}} \\ \underline{\mathbf{b}} \\ \underline{\mathbf{c}} \end{array}\right| = \left|\begin{array}{ccc} \underline{\mathbf{a}} & \underline{\mathbf{b}} & \underline{\mathbf{c}} \end{array}\right| = |A| \,,$$

Thus our definition of $\det A$ is equivalent to the earlier determinant notation $|A|$ of equation (7).

The form of the inverse obtained for $\mathbb{R}^{3\times3}$ in the geometric approach in §7 suggests the following theorem since it gives the following :

$$\begin{array}{rrcll} & (A^T)^{-1} & = & (A^{-1})^T & \text{(from §6)} \\ \Rightarrow & \dfrac{1}{\det A^T} \operatorname{adj} A^T & = & \left(\dfrac{1}{\det A} \operatorname{adj} A\right)^T & \text{(from §7)} \\ \Rightarrow & \dfrac{1}{\det A^T} \operatorname{adj} A^T & = & \dfrac{1}{\det A} (\operatorname{adj} A)^T & \end{array}$$

Theorem 18. For any $A \in \mathbb{F}^{3\times3}$, $\det A^T = \det A$ and $\operatorname{adj} A^T = (\operatorname{adj} A)^T$.

Proof. Let $A = \left[\begin{array}{ccc} \underline{\mathbf{a}} & \underline{\mathbf{b}} & \underline{\mathbf{c}} \end{array}\right] = \left[\begin{array}{ccc} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{array}\right]$, so $A^T = \left[\begin{array}{ccc} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{array}\right]$. Since $\det A = |A|$, the first equality follows immediately from the transpose property of the earlier determinant notation of §4.1.2. For the second equality :

$$\begin{array}{rcll} \operatorname{adj} A^T & = & \operatorname{adj} \left[\begin{array}{ccc} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{array}\right] & \\ & = & \left[\begin{array}{c} (a_2, b_2, c_2) \times (a_3, b_3, c_3) \\ (a_3, b_3, c_3) \times (a_1, b_1, c_1) \\ (a_1, b_1, c_1) \times (a_2, b_2, c_2) \end{array}\right], & \text{from Definition 14} \\ & = & \left[\begin{array}{ccc} b_2c_3 - b_3c_2 & -(a_2c_3 - a_3c_2) & a_2b_3 - a_3b_2 \\ b_3c_1 - b_1c_3 & -(a_3c_1 - a_1c_3) & a_3b_1 - a_1b_3 \\ b_1c_2 - b_2c_1 & -(a_1c_2 - a_2c_1) & a_1b_2 - a_2b_1 \end{array}\right] & (33) \end{array}$$

But

$$\begin{aligned}
\operatorname{adj} A &= \operatorname{adj} \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} \\
&= \begin{bmatrix} (b_1,b_2,b_3)\times(c_1,c_2,c_3) \\ (c_1,c_2,c_3)\times(a_1,a_2,a_3) \\ (a_1,a_2,a_3)\times(b_1,b_2,b_3) \end{bmatrix}, \qquad \text{from Definition 14} \\
&= \begin{bmatrix} b_2c_3-b_3c_2 & -(b_1c_3-b_3c_1) & b_1c_2-b_2c_1 \\ c_2a_3-c_3a_2 & -(c_1a_3-c_3a_1) & c_1a_2-c_2a_1 \\ a_2b_3-a_3b_2 & -(a_1b_3-a_3b_1) & a_1b_2-a_2b_1 \end{bmatrix} \qquad (34)
\end{aligned}$$

Comparing (33) and (34) we obtain $\operatorname{adj} A^T = (\operatorname{adj} A)^T$, as required.

QED

The following corollary gives the counterpart for a matrix specified row-wise to the adjugate and determinant for a matrix specified column-wise that was given in Definition 14.

Corollary. If $A \in \mathbb{F}^{3\times 3}$ is given by $A = \begin{bmatrix} \underline{\mathbf{a}} \\ \underline{\mathbf{b}} \\ \underline{\mathbf{c}} \end{bmatrix}$ then :

$$\begin{aligned}
\operatorname{adj} A &= \begin{bmatrix} \underline{\mathbf{b}}\times\underline{\mathbf{c}} & \underline{\mathbf{c}}\times\underline{\mathbf{a}} & \underline{\mathbf{a}}\times\underline{\mathbf{b}} \end{bmatrix}, \\
\text{and} \qquad \det A &= \underline{\mathbf{a}}\cdot(\underline{\mathbf{b}}\times\underline{\mathbf{c}})
\end{aligned}$$

□

Theorem 19. For any $A \in \mathbb{F}^{3\times 3}$, whether non-singular or singular, $(\operatorname{adj} A)A = (\det A)I$ and $A(\operatorname{adj} A) = (\det A)I$.

Proof. Taking $A = \begin{bmatrix} \underline{\mathbf{a}} & \underline{\mathbf{b}} & \underline{\mathbf{c}} \end{bmatrix}$, the first equality is immediate from Definition 14 and the properties of scalar and vector product on $\mathbb{F}^3$ in §9.3. For the second equality we have :

$$\begin{aligned}
& (\operatorname{adj} A^T)A^T = (\det A^T)I, && \text{from the first equality applied to } A^T \\
\Rightarrow\ & (\operatorname{adj} A)^T A^T = (\det A)I, && \text{from Theorem 18} \\
\Rightarrow\ & (A \operatorname{adj} A)^T = (\det A)I, && \text{from §6} \\
\Rightarrow\ & A(\operatorname{adj} A) = (\det A)I
\end{aligned}$$

QED

The following theorem is the counterpart of Theorem 11 for $\mathbb{R}^3$. Notice however that the method of proof is different and does not rely upon any geometric notions and is a purely algebraic 'a priori' proof.

Theorem 20. Let $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ be three arbitrary vectors in $\mathbb{F}^3$ and let A be the matrix $\begin{bmatrix} \underline{\mathbf{a}} & \underline{\mathbf{b}} & \underline{\mathbf{c}} \end{bmatrix}$. Then the following three statements are equivalent :

(a) A is invertible

(b) $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ are LI

(c) $\det A$ is non-zero

and in this case the inverse of A is given by :

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A$$

Proof. (a) ⇒ (b). Suppose (b) were false, ie. $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ are LD. Then one of these vectors is a linear combination of the other two, WLOG suppose it is $\underline{\mathbf{c}}$. Then if we right multiply A by a matrix B then the matrix $AB = \begin{bmatrix} \underline{\mathbf{d}} & \underline{\mathbf{e}} & \underline{\mathbf{f}} \end{bmatrix}$ produced has columns $\underline{\mathbf{d}}, \underline{\mathbf{e}}, \underline{\mathbf{f}} \in \mathbb{F}^3$ which are each linear combinations of $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$. Thus $\underline{\mathbf{d}}, \underline{\mathbf{e}}, \underline{\mathbf{f}} \in \operatorname{sp}(\underline{\mathbf{a}}, \underline{\mathbf{b}})$, and so by Theorem 14, $\underline{\mathbf{d}}\cdot(\underline{\mathbf{e}}\times\underline{\mathbf{f}}) = 0$. Thus it is impossible for $AB = I$, since $I = \begin{bmatrix} \underline{\mathbf{i}} & \underline{\mathbf{j}} & \underline{\mathbf{k}} \end{bmatrix}$ and $\underline{\mathbf{i}}\cdot(\underline{\mathbf{j}}\times\underline{\mathbf{k}}) = 1$. Hence A cannot have an inverse matrix. Thus ¬ (b) ⇒ ¬ (a), as required.

(b) ⇒ (c). This is established by Theorem 13 above.

(c) ⇒ (a). In Theorem 19, dividing both equalities through by $\det A$ gives a left and right inverse of A which is the required matrix.

QED

10.1 3 x 3 Product Formulae

Consider a product AB of two matrices $A, B \in \mathbb{F}^{3\times 3}$. From §6, if A and B are invertible then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$, thus from Theorem 20 :

$$\frac{1}{\det AB}\operatorname{adj} AB = (AB)^{-1} = B^{-1}A^{-1} = \frac{1}{\det B}\operatorname{adj} B \cdot \frac{1}{\det A}\operatorname{adj} A = \frac{1}{\det A \cdot \det B}\operatorname{adj} B \cdot \operatorname{adj} A \tag{35}$$

This suggests that it might generally be true that $\det AB = \det A \cdot \det B$ and $\operatorname{adj} AB = \operatorname{adj} B \cdot \operatorname{adj} A$, and we now prove this is the case in the results below.

Theorem 21. Given $A, B \in \mathbb{F}^{3\times 3}$, AB is invertible iff A and B are invertible.

Proof. The ⇐ direction is clear from inverses of matrix products in $\mathbb{F}^{n\times n}$ generally (§6).

⇒. We show $\neg q \Rightarrow \neg p$. Firstly suppose A is singular. Then by Theorem 20 the columns of A are LD vectors in $\mathbb{F}^3$, so one of the columns can be expressed as a linear combination of the other two, call the latter $\underline{\mathbf{d}}$, $\underline{\mathbf{e}}$. Then, forming the product AB the three columns produced, $\underline{\mathbf{a}}$, $\underline{\mathbf{b}}$, $\underline{\mathbf{c}}$ say, are each linear combinations of $\underline{\mathbf{d}}$ and $\underline{\mathbf{e}}$, thus $\underline{\mathbf{a}}, \underline{\mathbf{b}}, \underline{\mathbf{c}} \in \operatorname{sp}(\underline{\mathbf{d}}, \underline{\mathbf{e}})$, and hence by Theorem 13, $\det AB = \underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) = 0$ and hence from Theorem 20, AB is singular. Thus right-multiplying a singular matrix by any matrix (even an invertible matrix) results in a singular matrix. Consider the case now where B is singular. Then B^T is singular, so right-multiplying by A^T, B^TA^T is singular, ie. $(AB)^T$ is singular, and thus from §6 AB is singular, as required.

QED

Corollary. Given $A, B \in \mathbb{F}^{3\times 3}$, if B is a one-sided inverse of A then A is invertible with inverse B.

Proof. If B is a right inverse then $AB = I$ is invertible and so by Theorem 21, A must be invertible. Left multiplying the equation $AB = I$ by A^{-1} then gives $B = A^{-1}$. The proof is similar for B a left inverse.

QED

Theorem 22. Given $A, B \in \mathbb{F}^{3\times 3}$, $\operatorname{adj} AB = \operatorname{adj} B \cdot \operatorname{adj} A$ and $\det AB = \det A \cdot \det B$.

Proof. Let $A = \begin{bmatrix} \underline{\mathbf{a}} \\ \underline{\mathbf{b}} \\ \underline{\mathbf{c}} \end{bmatrix}$ and $B = \begin{bmatrix} \underline{\mathbf{d}} & \underline{\mathbf{e}} & \underline{\mathbf{f}} \end{bmatrix}$.

Then

$$AB = \begin{bmatrix} \underline{\mathbf{a}}\cdot\underline{\mathbf{d}} & \underline{\mathbf{a}}\cdot\underline{\mathbf{e}} & \underline{\mathbf{a}}\cdot\underline{\mathbf{f}} \\ \underline{\mathbf{b}}\cdot\underline{\mathbf{d}} & \underline{\mathbf{b}}\cdot\underline{\mathbf{e}} & \underline{\mathbf{b}}\cdot\underline{\mathbf{f}} \\ \underline{\mathbf{c}}\cdot\underline{\mathbf{d}} & \underline{\mathbf{c}}\cdot\underline{\mathbf{e}} & \underline{\mathbf{c}}\cdot\underline{\mathbf{f}} \end{bmatrix}$$

and so from Definition 14, taking the cross products of the columns of AB and placing them in the rows of $\operatorname{adj} AB$:

$$\operatorname{adj} AB = \begin{bmatrix} (\underline{\mathbf{b}}\cdot\underline{\mathbf{e}})(\underline{\mathbf{c}}\cdot\underline{\mathbf{f}}) - (\underline{\mathbf{c}}\cdot\underline{\mathbf{e}})(\underline{\mathbf{b}}\cdot\underline{\mathbf{f}}) & (\underline{\mathbf{c}}\cdot\underline{\mathbf{e}})(\underline{\mathbf{a}}\cdot\underline{\mathbf{f}}) - (\underline{\mathbf{a}}\cdot\underline{\mathbf{e}})(\underline{\mathbf{c}}\cdot\underline{\mathbf{f}}) & (\underline{\mathbf{a}}\cdot\underline{\mathbf{e}})(\underline{\mathbf{b}}\cdot\underline{\mathbf{f}}) - (\underline{\mathbf{b}}\cdot\underline{\mathbf{e}})(\underline{\mathbf{a}}\cdot\underline{\mathbf{f}}) \\ (\underline{\mathbf{b}}\cdot\underline{\mathbf{f}})(\underline{\mathbf{c}}\cdot\underline{\mathbf{d}}) - (\underline{\mathbf{c}}\cdot\underline{\mathbf{f}})(\underline{\mathbf{b}}\cdot\underline{\mathbf{d}}) & (\underline{\mathbf{c}}\cdot\underline{\mathbf{f}})(\underline{\mathbf{a}}\cdot\underline{\mathbf{d}}) - (\underline{\mathbf{a}}\cdot\underline{\mathbf{f}})(\underline{\mathbf{c}}\cdot\underline{\mathbf{d}}) & (\underline{\mathbf{a}}\cdot\underline{\mathbf{f}})(\underline{\mathbf{b}}\cdot\underline{\mathbf{d}}) - (\underline{\mathbf{b}}\cdot\underline{\mathbf{f}})(\underline{\mathbf{a}}\cdot\underline{\mathbf{d}}) \\ (\underline{\mathbf{b}}\cdot\underline{\mathbf{d}})(\underline{\mathbf{c}}\cdot\underline{\mathbf{e}}) - (\underline{\mathbf{c}}\cdot\underline{\mathbf{d}})(\underline{\mathbf{b}}\cdot\underline{\mathbf{e}}) & (\underline{\mathbf{c}}\cdot\underline{\mathbf{d}})(\underline{\mathbf{a}}\cdot\underline{\mathbf{e}}) - (\underline{\mathbf{a}}\cdot\underline{\mathbf{d}})(\underline{\mathbf{c}}\cdot\underline{\mathbf{e}}) & (\underline{\mathbf{a}}\cdot\underline{\mathbf{d}})(\underline{\mathbf{b}}\cdot\underline{\mathbf{e}}) - (\underline{\mathbf{b}}\cdot\underline{\mathbf{d}})(\underline{\mathbf{a}}\cdot\underline{\mathbf{e}}) \end{bmatrix}$$

(by symmetry to go from row 1 to row 2 we just replace $\underline{\mathbf{e}}$, $\underline{\mathbf{f}}$ with $\underline{\mathbf{f}}$, $\underline{\mathbf{d}}$, and to go from row 2 to row 3 we just replace $\underline{\mathbf{f}}$, $\underline{\mathbf{d}}$ with $\underline{\mathbf{d}}$, $\underline{\mathbf{e}}$).

Thus from Theorem 17

$$\operatorname{adj} AB = \begin{bmatrix} (\underline{\mathbf{b}}\times\underline{\mathbf{c}})\cdot(\underline{\mathbf{e}}\times\underline{\mathbf{f}}) & (\underline{\mathbf{c}}\times\underline{\mathbf{a}})\cdot(\underline{\mathbf{e}}\times\underline{\mathbf{f}}) & (\underline{\mathbf{a}}\times\underline{\mathbf{b}})\cdot(\underline{\mathbf{e}}\times\underline{\mathbf{f}}) \\ (\underline{\mathbf{b}}\times\underline{\mathbf{c}})\cdot(\underline{\mathbf{f}}\times\underline{\mathbf{d}}) & (\underline{\mathbf{c}}\times\underline{\mathbf{a}})\cdot(\underline{\mathbf{f}}\times\underline{\mathbf{d}}) & (\underline{\mathbf{a}}\times\underline{\mathbf{b}})\cdot(\underline{\mathbf{f}}\times\underline{\mathbf{d}}) \\ (\underline{\mathbf{b}}\times\underline{\mathbf{c}})\cdot(\underline{\mathbf{d}}\times\underline{\mathbf{e}}) & (\underline{\mathbf{c}}\times\underline{\mathbf{a}})\cdot(\underline{\mathbf{d}}\times\underline{\mathbf{e}}) & (\underline{\mathbf{a}}\times\underline{\mathbf{b}})\cdot(\underline{\mathbf{d}}\times\underline{\mathbf{e}}) \end{bmatrix}$$

(again to go from row 1 to row 2 we just replace $\underline{\mathbf{e}}$, $\underline{\mathbf{f}}$ with $\underline{\mathbf{f}}$, $\underline{\mathbf{d}}$, and to go from row 2 to row 3 we just replace $\underline{\mathbf{f}}$, $\underline{\mathbf{d}}$ with $\underline{\mathbf{d}}$, $\underline{\mathbf{e}}$).

Thus

$$\begin{aligned} \text{adj}\, AB &= \begin{bmatrix} \underline{\mathbf{e}} \times \underline{\mathbf{f}} \\ \underline{\mathbf{f}} \times \underline{\mathbf{d}} \\ \underline{\mathbf{d}} \times \underline{\mathbf{e}} \end{bmatrix} \begin{bmatrix} \underline{\mathbf{b}} \times \underline{\mathbf{c}} & \underline{\mathbf{c}} \times \underline{\mathbf{a}} & \underline{\mathbf{a}} \times \underline{\mathbf{b}} \end{bmatrix} \\ \Rightarrow \qquad \text{adj}\, AB &= \text{adj}\, B \cdot \text{adj}\, A \end{aligned} \tag{36}$$

as required.

If AB is invertible then from Theorem 21, A and B are invertible so $\det AB, \det A, \det B$ are all non-zero and equation (35) applies. But by Theorems 19 and 20, adj AB cannot be the zero matrix if AB is invertible and thus from equations (35) and (36) we obtain :

$$\det AB = \det A \cdot \det B \tag{37}$$

as required.

In the case where AB is singular then from Theorem 21, either A or B is singular, so that either $\det A$ or $\det B$ is zero, so again (37) applies.

QED

11 3 x 3 Matrix Inverse In Terms Of Minors, Cofactors, And Leibniz And Laplace Formulae

This section applies to matrices in $\mathbb{F}^{3\times 3}$, where $\mathbb{F}$ is a general field. We re-express the matrix inverse in a way that naturally leads to a conjecture for the form of the general $n \times n$ matrix inverse over $\mathbb{F}$ (§12).

Definition 15. Given a matrix $A \in \mathbb{F}^{3\times 3}$ and $i, j \in \{1, 2, 3\}$:

(a) the $(i, j)^{\text{th}}$ minor matrix M_{ij} of A is the 2×2 matrix obtained from A by omitting row i and column j

(b) the $(i, j)^{\text{th}}$ minor m_{ij} of A is the determinant of the minor matrix M_{ij} (where the 2×2 determinant is as defined in §4.1.2)

(c) the $(i, j)^{\text{th}}$ cofactor s_{ij} of A is the signed minor $(-1)^{i+j}\, m_{ij}$

(d) the cofactor matrix CF(A) of A is the matrix $[s_{ij}] \in \mathbb{F}^{3\times 3}$ of cofactors of A

□

11.1 3 x 3 Adjugate In Terms Of Minors And Cofactors

Theorem 23. Given a matrix $A \in \mathbb{F}^{3\times 3}$, adj A is the transpose of the cofactor matrix of A.

Proof. Let $A = \begin{bmatrix} \underline{\mathbf{a}} & \underline{\mathbf{b}} & \underline{\mathbf{c}} \end{bmatrix} = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix}$, so by Definition 14, adj $A = \begin{bmatrix} \underline{\mathbf{b}} \times \underline{\mathbf{c}} \\ \underline{\mathbf{c}} \times \underline{\mathbf{a}} \\ \underline{\mathbf{a}} \times \underline{\mathbf{b}} \end{bmatrix}$. Then :

$$\begin{aligned} \text{adj}\, A &= \begin{bmatrix} b_2c_3 - b_3c_2 & b_3c_1 - b_1c_3 & b_1c_2 - b_2c_1 \\ c_2a_3 - c_3a_2 & c_3a_1 - c_1a_3 & c_1a_2 - c_2a_1 \\ a_2b_3 - a_3b_2 & a_3b_1 - a_1b_3 & a_1b_2 - a_2b_1 \end{bmatrix} \\ &= \begin{bmatrix} m_{11} & -m_{21} & m_{31} \\ -m_{12} & m_{22} & -m_{32} \\ m_{13} & -m_{23} & m_{33} \end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} s_{11} & s_{21} & s_{31} \\ s_{12} & s_{22} & s_{32} \\ s_{13} & s_{23} & s_{33} \end{bmatrix}$$

$$= \text{transpose of CF}(A).$$

QED

11.2 3 x 3 Determinant In Leibniz Form

In the Leibniz row form below, in every product of 3 elements of A we select a different column for each row, and in the Leibniz column form a different row for each column.

Theorem 24 (Leibniz Formulae For 3×3 Determinant). If $A = [a_{ij}] \in \mathbb{F}^{3\times 3}$, then :

$$\det A = \sum_{\sigma \in S_3} \text{sgn}(\sigma) \prod_{i=1}^{3} a_{i,\sigma(i)} \qquad \text{(Leibniz row form)}$$

$$\text{and} \qquad \det A = \sum_{\tau \in S_3} \text{sgn}(\tau) \prod_{j=1}^{3} a_{\tau(j),j} \qquad \text{(Leibniz column form)}$$

where S_3 is the symmetric group of degree 3 ([11, p71], [7, p9], [28]), ie. the set of all 6 possible bijections or 'permutations' on the set $[3] = \{1, 2, 3\}$, and sgn is the permutation signature function (Appendix A, [43, §10]).

Proof. We have :

$$\begin{aligned} \det A &= \det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \\ &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}, \quad \text{using the notation of §4.1.2} \\ &= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} \end{aligned} \tag{38}$$

Note the row indices are in order 123 in each of the six terms in (38). Table 2 below compares the signs of the terms in (38) with the signatures of the permutation functions σ that select the three column indices for each term (see [43, §11] for how signatures can be calculated) :

Table 2: 3×3 Determinant Term Signs And Column Permutation Signatures

	Column selection function σ for rows 1, 2, 3 of A	Sign of term	Signature sgn(σ)
Term 1	1, 2, 3	+	+1
Term 2	1, 3, 2	–	-1
Term 3	2, 1, 3	–	-1
Term 4	2, 3, 1	+	+1
Term 5	3, 1, 2	+	+1
Term 6	3, 2, 1	–	-1

From the table we see all the term signs equal the signatures of the respective σ functions, and hence the RHS of (38) equals the required Leibniz row form.

If we rearrange (38) as follows, to put the column indices in order 123 in each term :

$$\det A = a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33} + a_{31}a_{12}a_{23} + a_{21}a_{32}a_{13} - a_{31}a_{22}a_{13} \tag{39}$$

then the comparison of the signs of the six terms in (39) with the signatures of the corresponding permutation functions τ that select the three row indices in each term is as shown in Table 3 below :

Table 3: 3×3 Determinant Term Signs And Row Permutation Signatures

	Row selection function τ for columns 1, 2, 3 of A	Sign of term	Signature $\text{sgn}(\tau)$
Term 1	1, 2, 3	+	+1
Term 2	1, 3, 2	–	-1
Term 3	2, 1, 3	–	-1
Term 4	3, 1, 2	+	+1
Term 5	2, 3, 1	+	+1
Term 6	3, 2, 1	–	-1

Again we see all the term signs equal the signatures of the respective τ functions, and hence the RHS of (39) equals the required Leibniz column form.

QED

11.3 3 x 3 Determinant In Laplace Form

Theorem 25 (Laplace Formulae For 3×3 Determinant). If $A = [a_{ij}] \in \mathbb{F}^{3\times 3}$, then :

$$\det A = \sum_{j=1}^{3}(-1)^{i+j}a_{ij}m_{ij} \quad \text{(Laplace expansion along the } i^{\text{th}} \text{ row, } i \in \{1,2,3\})$$

$$\text{and} \qquad \det A = \sum_{i=1}^{3}(-1)^{i+j}a_{ij}m_{ij} \quad \text{(Laplace expansion down the } j^{\text{th}} \text{ column, } j \in \{1,2,3\})$$

Proof. We have :

$$\begin{aligned}
\det A &= \det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \\
&= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}, \quad \text{using the notation of §4.1.2} \\
&= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\
&= a_{11}m_{11} - a_{12}m_{12} + a_{13}m_{13} \\
&= \text{Laplace expansion along the } 1^{\text{st}} \text{ row}
\end{aligned}$$

If we define $\underline{\mathbf{a}}, \underline{\mathbf{b}}, \underline{\mathbf{c}}$ to be the rows of A, then we have :

$$\begin{aligned}
\det A &= \det \begin{bmatrix} \underline{\mathbf{a}} \\ \underline{\mathbf{b}} \\ \underline{\mathbf{c}} \end{bmatrix} \\
&= \underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}), \quad \text{from the corollary to Theorem 18} \\
&= \underline{\mathbf{b}} \cdot (\underline{\mathbf{c}} \times \underline{\mathbf{a}}), \quad \text{from Theorem 15} \\
&= \det \begin{bmatrix} \underline{\mathbf{b}} \\ \underline{\mathbf{c}} \\ \underline{\mathbf{a}} \end{bmatrix}, \quad \text{from the corollary to Theorem 18} \\
&= \det \begin{bmatrix} a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{11} & a_{12} & a_{13} \end{bmatrix}, \quad \text{interchanging the rows of } A \\
&= \begin{vmatrix} a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{11} & a_{12} & a_{13} \end{vmatrix}, \quad \text{using the notation of §4.1.2} \\
&= a_{21} \begin{vmatrix} a_{32} & a_{33} \\ a_{12} & a_{13} \end{vmatrix} - a_{22} \begin{vmatrix} a_{31} & a_{33} \\ a_{11} & a_{13} \end{vmatrix} + a_{23} \begin{vmatrix} a_{31} & a_{32} \\ a_{11} & a_{12} \end{vmatrix} \\
&= -a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{22} \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} - a_{23} \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} \\
&= -a_{21} m_{21} + a_{22} m_{22} - a_{23} m_{23} \\
&= \text{Laplace expansion along the } 2^{\text{nd}} \text{ row.}
\end{aligned}$$

Using the identity $\underline{\mathbf{a}} \cdot (\underline{\mathbf{b}} \times \underline{\mathbf{c}}) = \underline{\mathbf{c}} \cdot (\underline{\mathbf{a}} \times \underline{\mathbf{b}})$, from Theorem 15, we can similarly prove $\det A$ equals the Laplace expansion along the 3^{rd} row.

From Theorem 18 we have :

$$\det A = \det A^T = \det \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix},$$

and so from the Laplace expansion of $\det A^T$ along the 1^{st} row :

$$\begin{aligned}
\det A &= a_{11} \begin{vmatrix} a_{22} & a_{32} \\ a_{23} & a_{33} \end{vmatrix} - a_{21} \begin{vmatrix} a_{12} & a_{32} \\ a_{13} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} a_{12} & a_{22} \\ a_{13} & a_{23} \end{vmatrix} \\
&= a_{11} m_{11} - a_{21} m_{21} + a_{31} m_{31} \\
&= \text{Laplace expansion of } \det A \text{ down the } 1^{\text{st}} \text{ column,}
\end{aligned}$$

By taking the Laplace expansions of $\det A^T$ along the 2^{nd} and 3^{rd} rows we similarly obtain the Laplace expansions of $\det A$ down the 2^{nd} and 3^{rd} columns, respectively.

QED

12 n x n Matrix Inverse

In §11 above we obtained algebraic expressions for the determinant and adjugate (and hence inverse) of a 3×3 matrix over a general field $\mathbb{F}$ which are in a form which is readily extendable to an $n \times n$ matrix over $\mathbb{F}$ for all $n \geq 3$. In this section we formally define these extensions and prove that the counterparts of Theorems 11 and 20 for the 3×3 case hold for an $n \times n$ matrix over $\mathbb{F}$, ie. with the definitions that we make below of $\det A$ and $\text{adj}\, A$ for $A \in \mathbb{F}^{n \times n}$ we will prove (Theorem 39) that : A is invertible $\iff$ the columns of A are LI $\iff$ $\det A \neq 0$, and that in this case the inverse of A is given by $A^{-1} = \dfrac{1}{\det A} \text{adj}\, A$.

All the proofs in this section stand independently of the results above for the 3×3 case, no inductive assumption is required, so they fully cover all cases of $n \geq 3$ for a general field $\mathbb{F}$. However the ideas in this section have their origin in the patterns observed in the 3×3 case. The proofs in this section in fact also apply to the 2×2 case, taking the 1×1 determinant to be the single value in the 1×1 matrix.

12.1 n x n Determinant

The following definition extends to the $n \times n$ case the definition of the Leibniz row and column forms for the 3×3 matrix made in Theorem 24 above.

Definition 16. If $A = [a_{ij}] \in \mathbb{F}^{n \times n}$, then :

$$\text{LR}(A) = \sum_{\sigma \in S_n} \text{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)} \qquad \text{(Leibniz row form for } A\text{)} \tag{40}$$

$$\text{LC}(A) = \sum_{\tau \in S_n} \text{sgn}(\tau) \prod_{j=1}^{n} a_{\tau(j),j} \qquad \text{(Leibniz column form for } A\text{)} \tag{41}$$

where S_n is the symmetric group of degree n ([11, p71], [7, p9], [28]), ie. the set of all $n!$ possible bijections or 'permutations' on the set $[n] = \{1, 2, \ldots, n\}$, and sgn is the permutation signature function (Appendix A, [43, §10]).

□

The following theorem shows that as in the 3×3 case the Leibniz row and column forms are equal.

Theorem 26. For any $A = [a_{ij}] \in \mathbb{F}^{n \times n}$, $\text{LR}(A) = \text{LC}(A)$.

Proof. For any $\sigma \in S_n$ the product $\prod_{i=1}^{n} a_{i,\sigma(i)}$ in $\text{LR}(A)$ may be written $\prod_{i=1}^{n} a_{\sigma^{-1}(i),\sigma(\sigma^{-1}(i))}$, since σ^{-1} is a bijection on $[n]$, so that we are just multiplying the same set of terms but in a different order. Thus $\prod_{i=1}^{n} a_{i,\sigma(i)} = \prod_{i=1}^{n} a_{\sigma^{-1}(i),i}$. Then by changing the dummy variable i in this product to j and using the fact that $\text{sgn}(\sigma^{-1}) = \text{sgn}(\sigma)$ ([43, §10]), we have from (40) :

$$\begin{aligned}
\text{LR}(A) &= \sum_{\sigma \in S_n} \text{sgn}(\sigma^{-1}) \prod_{j=1}^{n} a_{\sigma^{-1}(j),j} \\
&= \sum_{\sigma \in S_n} \text{sgn}(\sigma) \prod_{j=1}^{n} a_{\sigma(j),j}\,, \qquad \text{since } \sigma^{-1} \text{ ranges over } S_n \text{ as } \sigma \text{ ranges over } S_n \\
&= \sum_{\tau \in S_n} \text{sgn}(\tau) \prod_{j=1}^{n} a_{\tau(j),j}\,, \qquad \text{changing the dummy variable } \sigma \text{ to } \tau \\
&= \text{LC}(A)\,, \qquad \text{from (41).}
\end{aligned}$$

QED

We can now use the common value of the Leibniz row and column forms of A as a definition for $\det A$.

Definition 17. If $A = [a_{ij}] \in \mathbb{F}^{n \times n}$ then the determinant of A is defined to be the common value of $\text{LR}(A)$ and $\text{LC}(A)$,

denoted by :

$$\det A = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$$

□

From Theorem 24 this is compatible with Definition 14 of the 3×3 determinant over $\mathbb{F}$ we made earlier and with the vertical bar notations of §4.1.2. When $n = 2$, $\mathrm{LR}(A)$ in (40) evaluates to the 2×2 determinant $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$, as required.

Immediately we now have :

Theorem 27. For any $A = [a_{ij}] \in \mathbb{F}^{n\times n}$, $\det A^T = \det A$.

Proof. Let $A^T = [b_{ij}]$, so $b_{ij} = a_{ji}\ \forall\ i, j$. Then :

$$\det A^T = \mathrm{LR}(A^T) = \sum_{\sigma\in S_n} \mathrm{sgn}(\sigma) \prod_{i=1}^{n} b_{i,\sigma(i)} = \sum_{\sigma\in S_n} \mathrm{sgn}(\sigma) \prod_{i=1}^{n} a_{\sigma(i),i}$$

which with a change of the dummy variables σ, i to τ, j respectively gives $\mathrm{LC}(A)$, ie. $\det A$, as required.

QED

And $\det I = 1$, where I is the identity matrix in $\mathbb{F}^{n\times n}$, since in (40) the only term in the summation which is non-zero occurs when $\sigma \in S_n$ is the identity permutation on $[n]$, and this term is the product of the 1's down the diagonal.

Theorem 28. If $A \in \mathbb{F}^{n\times n}$ and if two rows (or two columns) of A are swapped then the determinant of A is multiplied by -1.

Proof. Suppose that rows k and l of A are swapped to produce $B = [b_{ij}]$. Then we have :

$$\begin{aligned} \det B &= \sum_{\sigma\in S_n} \mathrm{sgn}(\sigma) \prod_{i=1}^{n} b_{i,\sigma(i)}\ , \qquad \text{from the Leibniz row form (40) for } B \\ &= \sum_{\sigma\in S_n} \mathrm{sgn}(\sigma) \cdot b_{k,\sigma(k)} \cdot b_{l,\sigma(l)} \prod_{\substack{i=1,\\ i\notin\{k,l\}}}^{n} b_{i,\sigma(i)} \\ &= \sum_{\sigma\in S_n} \mathrm{sgn}(\sigma) \cdot a_{l,\sigma(k)} \cdot a_{k,\sigma(l)} \prod_{\substack{i=1,\\ i\notin\{k,l\}}}^{n} a_{i,\sigma(i)}\ , \qquad (42) \end{aligned}$$

since row k of B is row l of A and vice-versa, and all other rows of B are equal to those of A.

Now for each σ in S_n, define $\tau_\sigma \in S_n$ by :

$$\tau_\sigma(i) = \begin{cases} \sigma(i), & \text{for } i \notin \{k, l\}, \\ \sigma(l), & \text{for } i = k, \\ \sigma(k), & \text{for } i = l. \end{cases}$$

Then $\tau_\sigma = \sigma\rho$, where $\rho \in S_n$ is defined by :

$$\rho(i) = \begin{cases} i, & \text{for } i \notin \{k, l\}, \\ l, & \text{for } i = k, \\ k, & \text{for } i = l. \end{cases}$$

As a permutation of n objects, ρ is a transposition (it interchanges the objects at positions k and l), and so $\mathrm{sgn}(\rho) = -1$. Thus $\mathrm{sgn}(\tau_\sigma) = \mathrm{sgn}(\sigma\rho) = \mathrm{sgn}(\sigma) \cdot \mathrm{sgn}(\rho) = -\mathrm{sgn}(\sigma)$. We can now rewrite equation (42) as :

$$\begin{aligned} \det B &= -\sum_{\sigma\in S_n} \mathrm{sgn}(\tau_\sigma) \cdot a_{l,\tau_\sigma(l)} \cdot a_{k,\tau_\sigma(k)} \prod_{\substack{i=1,\\ i\notin\{k,l\}}}^{n} a_{i,\tau_\sigma(i)} \\ &= -\sum_{\sigma\in S_n} \mathrm{sgn}(\tau_\sigma) \prod_{i=1}^{n} a_{i,\tau_\sigma(i)} \qquad (43) \end{aligned}$$

But the mapping $\sigma \mapsto \tau_\sigma$ is a bijection on the group S_n, and so iterating over all τ_σ for all $\sigma \in S_n$ in equation (43) is equivalent to just iterating over all $\sigma \in S_n$ (the terms in the summation are just added up in a different order), so that :

$$\det B \quad = \quad -\sum_{\sigma \in S_n} \text{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)},$$

and the summation on the RHS of this is the Leibniz row form (40) for A, ie. $\det A$, as required.

Consider now the case where columns k and l of A are swapped to produce $B = [b_{ij}]$. Then B^T is produced from A^T by swapping rows k and l of A^T, so by the above $\det B^T = -\det A^T$ and then by Theorem 27, $\det B = -\det A$.

QED

Theorem 29. If $A \in \mathbb{F}^{n \times n}$ has two equal rows or has two equal columns then $\det A = 0$.

Proof. From Theorem 28 above if A has rows k and l equal then on swapping these rows, since this does not change the matrix, we have $\det A = -\det A$, and hence $2 \cdot \det A = 0$. Thus provided char $\mathbb{F} \neq 2$ (§8), so that $2 \neq 0$ in the field $\mathbb{F}$, we obtain $\det A = 0$. However if char $\mathbb{F} = 2$, ie. $\mathbb{F}$ is the prime field GF(2) of order 2, then $2 = 0$ (0 and 1 are the only elements in GF(2)), and so a separate proof is given below which covers this case.

As in the proof of Theorem 28 above, for every $\sigma \in S_n$ define $\tau_\sigma \in S_n$ to be the permutation which is identical to σ except that it swaps around the positions that positions k and l are mapped to, ie. it maps position k to position $\sigma(l)$ and position l to position $\sigma(k)$, but maps all other positions exactly as σ. Again as in Theorem 28 define ρ to be the permutation which swaps positions k and l, so that $\tau_\sigma = \sigma\rho$, $\text{sgn}(\rho) = -1$, and $\text{sgn}(\tau_\sigma) = \text{sgn}(\sigma) \cdot \text{sgn}(\rho) = -\text{sgn}(\sigma)$.

Define $T : S_n \mapsto S_n$ by $T(\sigma) = \tau_\sigma$, ie. $T(\sigma) = \sigma\rho$. Then since S_n is a group, T is a bijection on S_n. For each $\sigma \in S_n$ define the T-orbit of σ to be the set $o(\sigma) = \{T^n(\sigma) : n \in \mathbb{W}\} \subseteq S_n$, ($T^0$ being the identity mapping on S_n). Since $T(\sigma) = \sigma\rho$ and $T^2(\sigma) = T(\sigma\rho) = \sigma\rho^2 = \sigma$ we have $o(\sigma) = o(\tau_\sigma) = \{\sigma, \tau_\sigma\}$ and $|o(\sigma)| = 2, \forall\ \sigma \in S_n$. Let $\mathcal{P}$ be the collection $\{o(\sigma) : \sigma \in S_n\}$ of subsets of S_n. Then $f \in o \in \mathcal{P} \Rightarrow o = o(f)$. So existence of an $h \in o(f) \cap o(g) \Rightarrow o(f) = o(h)$ and $o(g) = o(h)$, and hence $o(f) = o(g)$. Thus if $o_1, o_2 \in \mathcal{P}$ and $o_1 \cap o_2 \neq \emptyset$ then $o_1 = o_2$, so $\mathcal{P}$ is a partition of the set S_n, into $\frac{1}{2}n!$ subsets each consisting of 2 elements. Now we can write :

$$\begin{aligned} \det A \quad &= \quad \sum_{\sigma \in S_n} \text{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)} \\ &= \quad \sum_{o \in \mathcal{P}} \sum_{\upsilon \in o} \text{sgn}(\upsilon) \prod_{i=1}^{n} a_{i,\upsilon(i)} \end{aligned} \tag{44}$$

Let $o \in \mathcal{P}$ and consider the inner sum in equation (44) corresponding to o. Suppose $o = \{\phi, \tau_\phi\}$. Then :

$$\sum_{\upsilon \in o} \text{sgn}(\upsilon) \prod_{i=1}^{n} a_{i,\upsilon(i)} \quad = \quad \text{sgn}(\phi) \cdot a_{k,\phi(k)} \cdot a_{l,\phi(l)} \prod_{\substack{i=1, \\ i \notin \{k,l\}}}^{n} a_{i,\phi(i)} + \text{sgn}(\tau_\phi) \cdot a_{k,\tau_\phi(k)} \cdot a_{l,\tau_\phi(l)} \prod_{\substack{i=1, \\ i \notin \{k,l\}}}^{n} a_{i,\tau_\phi(i)} \tag{45}$$

Now since rows k and l of A are identical, and since $\tau_\phi(k) = \phi(l)$ and $\tau_\phi(l) = \phi(k)$, we have :

$$\begin{aligned} a_{k,\phi(k)} \quad &= \quad a_{l,\tau_\phi(l)} \\ \text{and} \qquad a_{l,\phi(l)} \quad &= \quad a_{k,\tau_\phi(k)}. \end{aligned}$$

But also we have $\tau_\phi(i) = \phi(i)$ for $i \notin \{k, l\}$, and $\text{sgn}(\tau_\phi) = -\text{sgn}(\phi)$, and so RHS of (45) is zero. Hence every inner sum in (44) is zero, and thus $\det A = 0$.

If A has two equal columns then A^T has two equal rows and so by the above $\det A^T = 0$. But from Theorem 27 $\det A = \det A^T$, so $\det A = 0$, as required.

QED

12.2 Determinant As A Multilinear Map

The determinant defines a function of n variables from $\mathbb{F}^n \times \cdots \times \mathbb{F}^n \mapsto \mathbb{F}$ by :

$$\det\,(\underline{\mathbf{r}}_1,\ \ldots,\ \underline{\mathbf{r}}_n) \quad = \quad \begin{vmatrix} \underline{\mathbf{r}}_1 \\ \vdots \\ \underline{\mathbf{r}}_n \end{vmatrix}$$

We show below that this is a multilinear map of n variables (aka n-linear map) [29], ie. for each $k \in [n]$ the mapping $f_k : \mathbb{F}^n \mapsto \mathbb{F}$ defined by allowing the k^{th} variable to vary whilst holding the other $n-1$ variables fixed, ie.

$$f_k(\underline{\mathbf{v}}) \quad = \quad \det\left(\underline{\mathbf{r}}_1,\ \ldots,\ \underline{\mathbf{v}},\ \ldots,\ \underline{\mathbf{r}}_n\right), \qquad (\underline{\mathbf{v}} \text{ in the } k^{\text{th}} \text{ position})$$

is a linear mapping, ie. $f_k(\underline{\mathbf{v}} + \underline{\mathbf{w}}) = f_k(\underline{\mathbf{v}}) + f_k(\underline{\mathbf{w}})$ and $f_k(\lambda\underline{\mathbf{v}}) = \lambda\, f_k(\underline{\mathbf{v}})$.

If we denote the matrix formed by the n row vectors by $A = [a_{ij}]$ and allow the k^{th} row to vary as a variable $\underline{\mathbf{v}} = (v_1, \ldots, v_n)$, then from the Leibniz row form (40), if $\underline{\mathbf{u}} = \underline{\mathbf{v}} + \underline{\mathbf{w}}$:

$$\begin{aligned}
f_k(\underline{\mathbf{v}}) &= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \cdot v_{\sigma(k)} \prod_{\substack{i=1,\\ i\neq k}}^{n} a_{i,\sigma(i)}\ , \\
f_k(\underline{\mathbf{w}}) &= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \cdot w_{\sigma(k)} \prod_{\substack{i=1,\\ i\neq k}}^{n} a_{i,\sigma(i)}\ , \\
f_k(\underline{\mathbf{u}}) &= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \cdot u_{\sigma(k)} \prod_{\substack{i=1,\\ i\neq k}}^{n} a_{i,\sigma(i)}\ ,
\end{aligned}$$

where $u_{\sigma(k)} = v_{\sigma(k)} + w_{\sigma(k)}$ for every $\sigma \in S_n$, so that $f_k(\underline{\mathbf{u}}) = f_k(\underline{\mathbf{v}}) + f_k(\underline{\mathbf{w}})$. And, for $\lambda \in \mathbb{F}$ we have :

$$\begin{aligned}
f_k(\lambda\underline{\mathbf{v}}) &= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \cdot \lambda v_{\sigma(k)} \prod_{\substack{i=1,\\ i\neq k}}^{n} a_{i,\sigma(i)} \\
&= \lambda\, f_k(\underline{\mathbf{v}}).
\end{aligned}$$

Similarly defining a function of n variables from $\mathbb{F}^n \times \cdots \times \mathbb{F}^n \mapsto \mathbb{F}$ column-wise by:

$$\det\left(\underline{\mathbf{c}}_1,\ \ldots,\ \underline{\mathbf{c}}_n\right) \quad = \quad \left|\ \underline{\mathbf{c}}_1 \quad \cdots \quad \underline{\mathbf{c}}_n\ \right|$$

produces a multilinear map of n variables.

12.3 n x n Minors And The Adjugate

Having made a definition of the determinant of an arbitrary dimension $n \times n$ matrix over $\mathbb{F}$, using the Leibniz form, we can now give the following $n \times n$ counterpart to the above Definition 15 for the 3×3 case.

Definition 18. Given a matrix $A \in \mathbb{F}^{n\times n}$ and $i, j \in [n]$:

(a) the $(i,j)^{\text{th}}$ minor matrix M_{ij} of A is the $(n-1) \times (n-1)$ matrix obtained from A by omitting row i and column j

(b) the $(i,j)^{\text{th}}$ minor m_{ij} of A is the determinant of the $(n-1) \times (n-1)$ minor matrix M_{ij}

(c) the $(i,j)^{\text{th}}$ cofactor s_{ij} of A is the signed minor $(-1)^{i+j}\, m_{ij}$

(d) the cofactor matrix $\operatorname{CF}(A)$ of A is the matrix $[s_{ij}] \in \mathbb{F}^{n\times n}$ of cofactors of A

□

We now define the adjugate for the $n \times n$ case by generalizing from the 3×3 case. Definition 19 below corresponds with Theorem 23 above for the 3×3 adjugate, and is consistent with the result of that theorem.

Definition 19. Given a matrix $A \in \mathbb{F}^{n\times n}$ the adjugate $\operatorname{adj} A$ of A is defined to be the transpose of the cofactor matrix of A, ie. $\operatorname{adj} A = \operatorname{CF}(A)^T$.

□

The following result relates the minors of a matrix with the minors of its transpose.

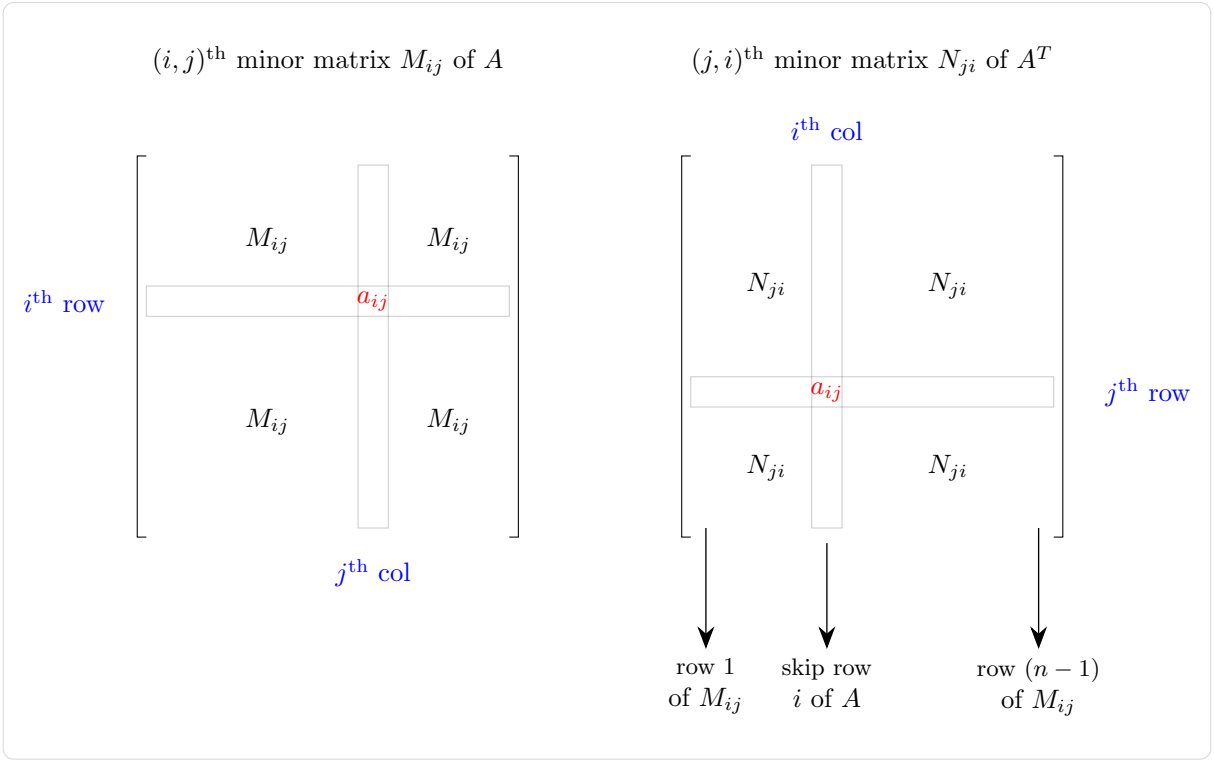
Theorem 30. If $A \in \mathbb{F}^{n\times n}$ then :

(a) the $(j,i)^{\text{th}}$ minor matrix N_{ji} of A^T is equal to M_{ij}^T, where M_{ij} is the $(i,j)^{\text{th}}$ minor matrix of A

(b) the $(j,i)^{\text{th}}$ minor n_{ji} of A^T is equal to the $(i,j)^{\text{th}}$ minor m_{ij} of A.

Proof. In Figure 16, which shows the four quadrants of M_{ij} and N_{ji} within A and A^T respectively, we can see that the broken rows of M_{ij} file in to become the broken columns of N_{ji}, so that we must have $N_{ji} = M_{ij}^T$. And then $n_{ji} = \det N_{ji} = \det M_{ij}^T = \det M_{ij} = m_{ij}$.

Figure 16: Minors Of Matrix And Transpose



QED

Theorem 31. For $A \in \mathbb{F}^{n\times n}$, $\operatorname{adj} A^T = (\operatorname{adj} A)^T$.

Proof.

$$\begin{aligned}
(i,j)^{\text{th}} \text{ of } \operatorname{adj} A^T &= (j,i)^{\text{th}} \text{ of } \operatorname{CF}(A^T)\,, \qquad \text{by Definition 19} \\
&= (-1)^{j+i}\, n_{ji}\,, \qquad \text{using notation of Theorem 30} \\
&= (-1)^{i+j}\, m_{ij}\,, \qquad \text{by Theorem 30(b)} \\
&= (i,j)^{\text{th}} \text{ of } \operatorname{CF}(A) \\
&= (j,i)^{\text{th}} \text{ of } \operatorname{adj} A\,, \qquad \text{by Definition 19} \\
&= (i,j)^{\text{th}} \text{ of } (\operatorname{adj} A)^T
\end{aligned}$$

QED

12.4 n x n Laplace Formulae

The following theorem is the $n \times n$ counterpart of Theorem 25 for the Laplace formulae in the 3×3 case.

Theorem 32 (Laplace Formulae For $n \times n$ Determinant). If $A = [a_{ij}] \in \mathbb{F}^{n\times n}$, then :

$$\det A = \sum_{j=1}^{n} (-1)^{i+j}\, a_{ij}\, m_{ij} \quad (\text{Laplace expansion along the } i^{\text{th}} \text{ row}, i \in [n]) \tag{46}$$

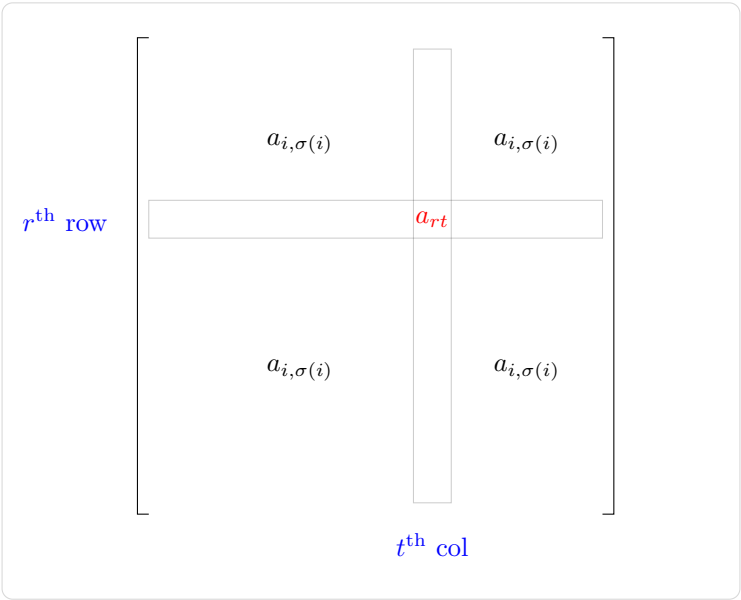
$$\text{and} \qquad \det A = \sum_{i=1}^{n} (-1)^{i+j}\, a_{ij}\, m_{ij} \quad (\text{Laplace expansion down the } j^{\text{th}} \text{ column}, j \in [n]) \tag{47}$$

Proof. First of all we consider the expansion along the r^{th} row, where $r \in [n]$. For each $t \in [n]$, define $S_n^{(t)}$ to be the set of all σ in S_n with $\sigma(r) = t$. Then S_n can be partitioned as $S_n = \bigcup_{t=1}^{n} S_n^{(t)}$, with $|S_n^{(t)}| = (n-1)!$ and $|S_n| = n!$. Let d_t be the contribution which $S_n^{(t)}$ makes to the Leibniz row form (40) for A. Then :

$$\begin{aligned} d_t &= \sum_{\sigma \in S_n^{(t)}} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)} \\ &= \sum_{\sigma \in S_n^{(t)}} \operatorname{sgn}(\sigma) \cdot a_{rt} \prod_{\substack{i=1, \\ i \neq r}}^{n} a_{i,\sigma(i)} \\ &= a_{rt} \left[\sum_{\sigma \in S_n^{(t)}} \operatorname{sgn}(\sigma) \prod_{\substack{i=1, \\ i \neq r}}^{n} a_{i,\sigma(i)} \right] \end{aligned}$$

in which for every $i \in [n] \setminus \{r\}$ we must have $\sigma(i) \in [n] \setminus \{t\}$, due to the bijective nature of the permutation functions σ, and hence every one of the terms $a_{i,\sigma(i)}$ is a term from the minor matrix $M_{rt} \in \mathbb{F}^{(n-1)\times(n-1)}$, as shown in Figure 17.

Figure 17: Elements of A selected by $\sigma \in S_n^{(t)}$



Now as in Theorem 6 of [43, §12], for each $\tau \in S_{n-1}$ define the extension permutation $\sigma_\tau \in S_n^{(t)}$ by :

$$\begin{aligned} \sigma_\tau(r) &= t\,, \\ \sigma_\tau \text{ on } [n] \setminus \{r\} &= \text{determined from } \tau \text{ by identifying } [n] \setminus \{r\} \text{ with } [n-1] \text{ in dom } \sigma_\tau \\ &\quad \text{and } [n] \setminus \{t\} \text{ with } [n-1] \text{ in ran } \sigma_\tau. \end{aligned}$$

So to evaluate σ_τ on $[n] \setminus \{r\}$ we :

(i) translate the argument into the range $[n-1]$,

(ii) apply τ to it,

(iii) translate the result from $[n-1]$ back into $[n] \setminus \{t\}$.

Then $\tau \mapsto \sigma_\tau$ is a 1–1 map from S_{n-1} to $S_n^{(t)}$, so σ ranging over $S_n^{(t)}$ is equivalent to σ_τ ranging over τ in S_{n-1}, and thus

we can write :

$$d_t \;=\; a_{rt}\left[\sum_{\tau\in S_{n-1}} \operatorname{sgn}(\sigma_\tau)\prod_{\substack{i=1,\\ i\neq r}}^{n} a_{i,\sigma_\tau(i)}\right].$$

Then if $M_{rt}=[b_{ij}]$, due to the relationship between τ and σ_τ, and observing the row-wise selections from Figure 17, we have :

$$\prod_{\substack{i=1,\\ i\neq r}}^{n} a_{i,\sigma_\tau(i)} \;=\; \prod_{i=1}^{n-1} b_{i,\tau(i)}\,, \qquad \text{for every } \tau\in S_{n-1}$$

$$\text{and hence,} \qquad d_t \;=\; a_{rt}\left[\sum_{\tau\in S_{n-1}} \operatorname{sgn}(\sigma_\tau)\prod_{i=1}^{n-1} b_{i,\tau(i)}\right].$$

But from Theorem 6 of [43, §12], we have $\operatorname{sgn}(\sigma_\tau)=(-1)^{r+t}\operatorname{sgn}(\tau)\ \forall\ \tau\in S_{n-1}$, and so :

$$\begin{aligned} d_t &= (-1)^{r+t}\,a_{rt}\left[\sum_{\tau\in S_{n-1}} \operatorname{sgn}(\tau)\prod_{i=1}^{n-1} b_{i,\tau(i)}\right] \\ &= (-1)^{r+t}\,a_{rt}\,m_{rt}\,, \qquad \text{from the Leibniz row form for } m_{rt}=\det M_{rt}. \end{aligned}$$

Then from the Leibniz row form (40) for A we have :

$$\det A=\sum_{t=1}^{n} d_t=\sum_{t=1}^{n}(-1)^{r+t}\,a_{rt}\,m_{rt}$$

which is the Laplace expansion (46) of $\det A$ along the r^{th} row, as required.

Now consider the Laplace expansion down the j^{th} column of A. From Theorem 27:

$$\begin{aligned} \det A &= \det A^T \\ &= \begin{vmatrix} a_{11} & a_{21} & \cdots\cdots & a_{n1} \\ \vdots & \vdots & & \vdots \\ a_{1j} & a_{2j} & \cdots\cdots & a_{nj} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots\cdots & a_{nn} \end{vmatrix} \\ &= \sum_{i=1}^{n}(-1)^{j+i}\,a_{ij}\,n_{ji}\,, \qquad \text{taking the Laplace expansion of } \det A^T \text{ along row } j \text{ (where} \\ &\qquad n_{ji} \text{ is the } (j,i)^{\text{th}} \text{ minor of } A^T) \\ &= \sum_{i=1}^{n}(-1)^{i+j}\,a_{ij}\,m_{ij}\,, \qquad \text{since from Theorem 30, } n_{ji}=m_{ij} \\ \Rightarrow \qquad \det A &= \text{Laplace expansion of } A \text{ down column } j, \text{ ie. (47).} \end{aligned}$$

QED

12.4.1 Laplace Off-Row And Off-Column Expansions

In this section we consider the notions of Laplace 'off-row' and 'off-column' expansions. These are best compared with the standard Laplace expansions by rewriting (46) and (47) in terms of the signed minors s_{ij} of Definition 18 as follows :

$$\det A \;=\; \sum_{j=1}^{n} a_{ij}\,s_{ij} \quad (\text{Laplace expansion along the } i^{\text{th}} \text{ row, } i\in[n]) \tag{48}$$

$$\text{and} \qquad \det A \;=\; \sum_{i=1}^{n} a_{ij}\,s_{ij} \quad (\text{Laplace expansion down the } j^{\text{th}} \text{ column, } j\in[n]) \tag{49}$$

In the off-row expansion we substitute the row coefficients a_{ij} from row i with row coefficients a_{kj} from a different row $k \neq i$. In the off-column expansion we substitute the column coefficients a_{ij} from column j with column coefficients a_{ik} from a different column $k \neq j$. In both cases we keep the signed minors s_{ij} the same as before — in (46) and (47) we could think of the sign as being attached to the minor m_{ij} rather than the row/column coefficient a_{ij}.

Theorem 33. If $A = [a_{ij}] \in \mathbb{F}^{n \times n}$, then :

$$\sum_{j=1}^{n} a_{kj}\, s_{ij} \quad = \quad 0 \quad \text{(Laplace 'off-row' expansion along the } i^{\text{th}} \text{ row, for } k \neq i) \tag{50}$$

$$\text{and} \qquad \sum_{i=1}^{n} a_{ik}\, s_{ij} \quad = \quad 0 \quad \text{(Laplace 'off-column' expansion down the } j^{\text{th}} \text{ column, for } k \neq j) \tag{51}$$

Proof. For $i \in [n]$ and $k \neq i$, let B be the matrix obtained from A by replacing row i of A with row k of A. Then rows i and k of B are equal and hence by Theorem 29, $\det B = 0$. Thus the Laplace expansion of B along its i^{th} row is zero, by Theorem 32. But since the i^{th} row of B is equal to the k^{th} row of A, and all the other rows of B are equal to those of A, so that the signed minors of B along row i are equal to the signed minors s_{ij} of A along row i, the summation in (50) is the Laplace expansion of B along its i^{th} row, ie. $\det B$, and hence is zero, as required.

For the second identity a similar argument applies, with B obtained from A by replacing column j of A with column k of A, so that columns j and k of B are equal and $\det B = 0$, and the summation in (51) is the Laplace expansion of B down its j^{th} column, and hence zero.

QED

12.5 Elementary Row Operations

Elementary row operations (ERO's) [10, p52], [31] allow conversion of an arbitrary matrix $A \in \mathbb{F}^{m \times n}$ into the following simplified forms which have a roughly upper triangular or staircase type shape [30], [10, p55] :

- Row echelon form (REF), in which :
  1. all the zero rows, if any, are at the bottom
  2. the leading entry (ie. the first or leftmost non-zero entry) of every non-zero row is strictly to the right of the leading entry of the row above it
- Reduced row echelon form (RREF), in which :
  1. it is in REF form
  2. the leading entry of each non-zero row is 1
  3. if a column contains a leading entry then every other entry in the column is zero

For example $\begin{bmatrix} 0 & 2 & 7 & 5 & 4 & 9 \\ 0 & 0 & 0 & 9 & 1 & 3 \\ 0 & 0 & 0 & 0 & 8 & 6 \\ 0 & 0 & 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ is in REF form and $\begin{bmatrix} 1 & 6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 9 \\ 0 & 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ is in RREF form, and the identity matrix I in $\mathbb{F}^{n \times n}$ is in RREF form.

Theorem 34. An RREF matrix in $\mathbb{F}^{n \times n}$ either has at least one zero row or it must be the identity matrix I in $\mathbb{F}^{n \times n}$.

Proof. If any leading 1 was more than one place to the right of the preceding leading 1 in the row above it, or if the leading 1 in the first row was not in the first column, then there would not be sufficiently many columns available to accommodate n leading 1's in the RREF, and hence at least one row must have no leading 1. Thus there must be at least one zero row.

Otherwise the leading 1 on row 1 must be in column 1 and every leading 1 thereafter must be exactly one place to the right of the preceding leading 1 in the row above it. Then the leading 1's must stairstep down the diagonal to reach the bottom right hand corner, and hence the RREF must be I.

QED

The three types of ERO's are :

(i) swapping two rows,

(ii) multiplying a row by a non-zero scalar,

(iii) adding a scalar multiple of one row to another row.

These can be achieved respectively by premultiplying the matrix in $\mathbb{F}^{m\times n}$ by the following 'elementary matrices' (EM's) in $\mathbb{F}^{m\times m}$, [10, p61] :

(i) for the ERO swapping rows k and l, the EM obtained by swapping rows k and l of the identity matrix $I \in \mathbb{F}^{m\times m}$,

(ii) for the ERO multiplying row k by scalar λ, the EM obtained by replacing the k^{th} unity on the diagonal of the identity matrix $I \in \mathbb{F}^{m\times m}$ with the value λ,

(iii) for the ERO adding scalar multiple λ of row l to row k ($k \neq l$), the EM obtained by adding a value λ at position (k, l) of the identity matrix $I \in \mathbb{F}^{m\times m}$.

Then performing an ERO on a matrix $A \in \mathbb{F}^{m\times n}$ produces the matrix EA, where E is the EM associated with the ERO. In particular, applying an ERO to $I \in \mathbb{F}^{m\times m}$ produces the EM $E \in \mathbb{F}^{m\times m}$ corresponding to the ERO.

A matrix $A \in \mathbb{F}^{m\times n}$ can be converted into REF form by the application of a sequence of ERO operations in a process called Gaussian elimination [32]. The steps are as follows :

1. Use row swaps to move all zero rows to the bottom.
2. Look for the first column c_1 which has a non-zero entry and perform a row swap if necessary to ensure that a non-zero entry is at the top of that column.
3. Perform a sequence of type (iii) ERO's to add appropriate multiples of row 1 to all the rows below row 1 so that all entries in column c_1 below the top one become zero. This is the elimination part of the process, with the leading entry of row 1 acting as the 'pivot' and row 1 as the 'pivotal row'.
4. Repeat steps 2 and 3 with the submatrix A_2 of A which comprises rows 2 through m.
5. Repeat steps 2 and 3 with the submatrix A_3 of A which comprises rows 3 through m.
6. Continue this process until the submatrix obtained comprises only zero rows, or there are no more submatrices left.

Note at each application of step 3 only rows of A below the pivotal row are affected.

To convert A into RREF form the following modification of the Gaussian elimination process can be applied, called Gauss–Jordan elimination, [32], [38] :

1. Immediately prior to carrying out step 3 apply a type (ii) ERO to change the value of the pivot to 1.
2. During the application of step 3 to a submatrix apply the elimination process to all the rows of A not just to the submatrix, ie. apply it to all the rows above the pivotal row as well as to those below it.

Each of the three types of ERO are reversible : type (i) reverses itself, type (ii) is reversed by multiplying row k by $1/\lambda$, and type (iii) is reversed by subtracting scalar multiple λ of row l from row k. Correspondingly the associated EM's are invertible :

(i) the EM which swaps rows k and l is its own inverse,

(ii) the inverse of the EM which multiplies row k by scalar λ is the matrix obtained from the identity matrix $I \in \mathbb{F}^{m\times m}$ by replacing the k^{th} unit on the diagonal with the value $1/\lambda$,

(iii) the inverse of the EM which adds scalar multiple λ of row l to row k is the matrix obtained from the identity matrix $I \in \mathbb{F}^{m\times m}$ by adding a value $-\lambda$ at position (k, l) .

We can define a relation '$\sim$' on $\mathbb{F}^{m\times n}$ called 'row equivalence' [10, p54] by : $A \sim B$ iff A can be transformed into B via a sequence of ERO's, ie. if there exists a sequence $E_1, \ldots, E_r$ of EM's such that $B = E_r \cdots E_1 A$. Then $A \sim B \Rightarrow A = E_1^{-1} \cdots E_r^{-1} B$, ie. $B \sim A$, so that $\sim$ is a symmetric relation. The relation $\sim$ is also clearly reflexive and transitive and hence is an equivalence relation [11, p36] on $\mathbb{F}^{m\times n}$. From the Gaussian and Gauss-Jordan elimination processes described above we now have (cf. [10, pp. 59, 61]) :

Theorem 35. Any matrix $A \in \mathbb{F}^{m\times n}$ is row equivalent to an REF and to an RREF.

QED

Theorem 36. If $A \in \mathbb{F}^{n\times n}$, the following statements are equivalent :

(a) A is invertible,

(b) $A \sim I$,

(c) A is a product of EM's.

Proof. (b) $\Rightarrow$ (a) & (c). Suppose $A \sim I$. Then $I \sim A$, and so there is a sequence of EM's, $E_1, \ldots, E_r$, such that $A = E_r \cdots E_1 I$, so (c) holds. Since any EM is invertible we also have $A^{-1} = E_1^{-1} \cdots E_r^{-1}$ and so (a) holds.

$\neg$ (b) $\Rightarrow \neg$ (a) & $\neg$ (c). By Theorems 34 and 35, $A \not\sim I$ implies $A \sim R$ for an RREF R with at least one zero row. Then $R = E_r \cdots E_1 A$, for some sequence of EM's $E_1, \ldots, E_r$. But R cannot have a right inverse since it has a zero row and therefore A must be singular, since the EM's are invertible, so (a) is false. Then (c) is also false, again since the EM's are invertible.

Hence (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (a).

QED

Definition 20.

(a) The row space of a matrix $A \in \mathbb{F}^{m \times n}$ given by :

$A = \begin{bmatrix} \underline{\mathbf{r}}_1 \\ \vdots \\ \underline{\mathbf{r}}_m \end{bmatrix}$ is the subspace of $\mathbb{F}^n$ spanned by the row vectors $\underline{\mathbf{r}}_1, \ldots, \underline{\mathbf{r}}_m$.

(b) The column space of a matrix $A \in \mathbb{F}^{m \times n}$ given by $A = \begin{bmatrix} \underline{\mathbf{c}}_1 & \cdots & \underline{\mathbf{c}}_n \end{bmatrix}$ is the subspace of $\mathbb{F}^m$ spanned by the column vectors $\underline{\mathbf{c}}_1, \ldots, \underline{\mathbf{c}}_n$.

□

Theorem 37.

(a) If $A \in \mathbb{F}^{m \times n}$, the application of an ERO to A does not alter the row space of A.

(b) If $A \in \mathbb{F}^{m \times n}$, the application of an ERO to A does not alter whether the row vectors of A are LI or LD.

(c) If $A \in \mathbb{F}^{n \times n}$, the application of an ERO to A multiplies $\det A$ by a factor of -1, λ, and $+1$, for types (i), (ii), and (iii) ERO respectively, where λ is the multiplication factor for the type (ii) ERO. Hence if E is an EM then

$$\det E = \begin{cases} -1, & \text{if } E \text{ is type (i)}, \\ \lambda, & \text{if } E \text{ is type (ii)}, \\ +1, & \text{if } E \text{ is type (iii)}. \end{cases}$$

Proof. (a) Let $A = \begin{bmatrix} \underline{\mathbf{r}}_1 \\ \vdots \\ \underline{\mathbf{r}}_m \end{bmatrix}$, and let S be the row space of A. Let B be the matrix A after an ERO has been applied to it and let T be the row space of B.

If the ERO is type (i) then two rows of A have been swapped and thus the span of the rows is unchanged, so $S = T$.

If the ERO is type (ii) then a row of A has been multiplied by a non-zero scalar λ and thus again straightforwardly the span of the rows is unchanged, so $S = T$.

Consider a type (iii) ERO, which adds a scalar multiple λ of row l to row k of A (assume $k < l$, the case $l < k$ follows similarly), so the rows vectors of B are : $\underline{\mathbf{r}}_1, \ldots, \underline{\mathbf{r}}_{k-1}, \underline{\mathbf{r}}_k + \lambda \underline{\mathbf{r}}_l, \underline{\mathbf{r}}_{k+1}, \ldots, \underline{\mathbf{r}}_n$. A linear combination of row vectors of B must clearly also be a linear combination of the row vectors of A, so we have $T \subseteq S$. Consider now a general member of S, $\underline{\mathbf{v}} = \mu_1 \underline{\mathbf{r}}_1 + \cdots + \mu_m \underline{\mathbf{r}}_m$ say. Then we can write $\underline{\mathbf{v}}$ as :

$$\underline{\mathbf{v}} = \mu_1 \underline{\mathbf{r}}_1 + \cdots + \mu_{k-1} \underline{\mathbf{r}}_{k-1} + \mu_k (\underline{\mathbf{r}}_k + \lambda \underline{\mathbf{r}}_l) + \mu_{k+1} \underline{\mathbf{r}}_{k+1} \cdots + \mu_{l-1} \underline{\mathbf{r}}_{l-1} + (\mu_l - \mu_k \lambda) \underline{\mathbf{r}}_l + \mu_{l+1} \underline{\mathbf{r}}_{l+1} \cdots + \mu_m \underline{\mathbf{r}}_m \tag{52}$$

so that $\underline{\mathbf{v}}$ is a linear combination of the row vectors of B, ie. $\underline{\mathbf{v}} \in T$. Thus $S \subseteq T$, and so $S = T$ as required.

(b) Let A and B be as in (a) above. It suffices to show the rows of A are LI iff the rows of B are LI.

For type (i) and (ii) ERO's this is straightforward.

Consider a type (iii) ERO, using the same notation as in (a) above. Suppose rows of B are LI and consider a linear combination of the rows of A equal to zero : $\underline{\mathbf{v}} = \mu_1\underline{\mathbf{r}}_1 + \cdots + \mu_m\underline{\mathbf{r}}_m = 0$ say. Then since $\underline{\mathbf{v}}$ equals the expression (52) in part (a) above, the LI of the rows of B implies $\mu_1 = \cdots = \mu_k = \cdots = \mu_{l-1} = 0$, $\mu_l - \mu_k\lambda = 0$, and $\mu_{l+1} = \cdots = \mu_m = 0$. Then $\mu_l = 0$ and so all of the coefficients μ_i of $\underline{\mathbf{v}}$ are zero. Thus rows of A are LI.

Conversely suppose rows of A are LI and consider a linear combination of the rows of B which is equal to zero :

$$\begin{aligned} & \mu_1\underline{\mathbf{r}}_1 + \cdots + \mu_{k-1}\underline{\mathbf{r}}_{k-1} + \mu_k(\underline{\mathbf{r}}_k + \lambda\underline{\mathbf{r}}_l) + \mu_{k+1}\underline{\mathbf{r}}_{k+1} + \cdots + \mu_l\underline{\mathbf{r}}_l + \cdots + \mu_m\underline{\mathbf{r}}_m &=& \ 0 \\ \Rightarrow \quad & \mu_1\underline{\mathbf{r}}_1 + \cdots + \mu_k\underline{\mathbf{r}}_k + \cdots + \mu_{l-1}\underline{\mathbf{r}}_{l-1} + (\mu_l + \mu_k\lambda)\underline{\mathbf{r}}_l + \mu_{l+1}\underline{\mathbf{r}}_{l+1}\cdots + \mu_m\underline{\mathbf{r}}_m &=& \ 0 \end{aligned}$$

Then LI of rows of A implies $\mu_1 = \cdots = \mu_k = \cdots = \mu_{l-1} = 0$, $\mu_l + \mu_k\lambda = 0$, and $\mu_{l+1} = \cdots = \mu_m = 0$. Then $\mu_l = 0$ and hence all the coefficients μ_i are zero. Thus rows of B are LI and so we have shown the rows of A are LI iff the rows of B are LI, as required.

(c) Let $A = \begin{bmatrix} \underline{\mathbf{r}}_1 \\ \vdots \\ \underline{\mathbf{r}}_n \end{bmatrix}$, and let B be the matrix A after an ERO has been applied to it.

If the ERO is type (i) then by Theorem 28, $\det A$ is multiplied by –1.

If the ERO is type (ii) then from the multilinearity of the determinant (§12.2), $\det A$ is multiplied by λ.

Consider a type (iii) ERO, which adds a scalar multiple λ of row l to row k of A, so the row vectors of B are : $\underline{\mathbf{r}}_1, \ldots, \underline{\mathbf{r}}_{k-1}, \underline{\mathbf{r}}_k + \lambda\underline{\mathbf{r}}_l, \underline{\mathbf{r}}_{k+1}, \ldots, \underline{\mathbf{r}}_n$. Then from the multilinearity of the determinant (§12.2) we have $\det B = \det A + \lambda \det C$, where C is the matrix obtained from A by replacing row k with row l. Since C then has two equal rows, by Theorem 29, $\det C = 0$ and hence $\det B = \det A$, as required.

The formula for $\det E$ then follows by applying the ERO to the matrix I, since $\det I = 1$.

QED

12.6 Main Theorem

Firstly we prove the following $n \times n$ counterpart of Theorem 19 for the 3×3 case.

Theorem 38. For any $A \in \mathbb{F}^{n\times n}$, whether non-singular or singular, $(\operatorname{adj} A)A = A(\operatorname{adj} A) = (\det A)I$.

Proof.

$$\begin{aligned} (j,k)^{\text{th}} \text{ of } (\operatorname{adj} A)A &= \text{product of } j^{\text{th}} \text{ row of adj } A \text{ and } k^{\text{th}} \text{ column of } A \\ &= \text{product of } j^{\text{th}} \text{ column of CF}(A) \text{ and } k^{\text{th}} \text{ column of } A, \text{ by Definition 19} \\ &= \sum_{i=1}^{n} a_{ik}\, s_{ij}\,, \qquad \text{from Definition 18(d)} \\ &= \begin{cases} \det A, & \text{if } k = j, \\ 0, & \text{if } k \neq j \end{cases} \\ & \quad \text{from the Laplace column expansion (49) and the Laplace off-column expansion (51)} \\ &= (j,k)^{\text{th}} \text{ of } (\det A)I. \end{aligned}$$

Hence $(\text{adj}\, A)A = (\det A)I$. To prove $A(\text{adj}\, A) = (\det A)I$:

$$\begin{aligned}
(k,i)^{\text{th}} \text{ of } A(\text{adj}\, A) &= \text{product of } k^{\text{th}} \text{ row of } A \text{ and } i^{\text{th}} \text{ column of adj}\, A \\
&= \text{product of } k^{\text{th}} \text{ row of } A \text{ and } i^{\text{th}} \text{ row of CF}(A), \text{ by Definition 19} \\
&= \sum_{j=1}^{n} a_{kj}\, s_{ij}\,, \qquad \text{from Definition 18(d)} \\
&= \begin{cases} \det A, & \text{if } k = i, \\ 0, & \text{if } k \neq i \end{cases} \\
&\quad \text{from the Laplace row expansion (48) and the Laplace off-row expansion (50)} \\
&= (k,i)^{\text{th}} \text{ of } (\det A)I.
\end{aligned}$$

Hence $A(\text{adj}\, A) = (\det A)I$.

QED

Theorem 39 (Main Theorem). Let $A \in \mathbb{F}^{n\times n}$. Then the following three statements are equivalent :

(a) A is invertible

(b) columns of A are LI vectors in $\mathbb{F}^n$

(c) $\det A$ is non-zero

and in this case the inverse of A is given by :

$$A^{-1} = \frac{1}{\det A}\, \text{adj}\, A.$$

Proof. (c) $\Rightarrow$ (a). This follows immediately from Theorem 38 by dividing through by $\det A$, and the two-sided inverse thus obtained is in the required form.

(a) $\Rightarrow$ (c). From Theorem 36 we have $A \sim I$, so there exists a sequence $E_1, \ldots, E_r$ of EM's such that $I = E_r \cdots E_1 A$. From Theorem 37(c), premultiplication of a matrix by an EM multiplies the determinant of the matrix by a factor, so that it could never change the determinant from zero to non-zero. Thus if $\det A = 0$ then this would imply $\det I = 0$, which is false, thus we must have $\det A \neq 0$.

(b) $\Rightarrow$ (a). Firstly suppose A was singular. From Theorems 34, 35, and 36 this means $A \sim R$ where R is an RREF with at least one zero row (hence R has rows which form an LD set in $\mathbb{F}^n$). Then there exists a sequence $E_1, \ldots, E_r$ of EM's such that $R = E_r \cdots E_1 A$. From Theorem 37(b), premultiplication of a matrix by an EM does not alter the LI/LD status of the row space of the matrix. Thus A must originally have had LD rows. Thus via the contrapositive we have shown that if A has LI rows then A must be invertible. Now consider the case (b) of A with LI columns. Then A^T has LI rows, and so A^T is invertible. Then from §6, A must be invertible, ie. (a) applies.

(c) $\Rightarrow$ (b). Consider the case of $\neg$ (b), ie. columns of A are LD. This implies one of the columns is a linear combination of the remaining $n-1$ columns. From the column-wise multilinearity of the determinant (§12.2) this means $\det A$ is a linear combination of $n-1$ determinants, each one of which is the determinant of a matrix which has a duplicate column. But from Theorem 29 a matrix with a duplicate column has determinant zero, hence we must have $\det A = 0$, ie. $\neg$ (c), as required.

Thus (a) and (c) are equivalent and these in turn are equivalent to (b).

QED

12.7 n x n Product Formulae

Consider a product AB of two matrices $A, B \in \mathbb{F}^{n\times n}$. From §6, if A and B are invertible then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$, thus from Theorem 39 :

$$\frac{1}{\det AB}\, \text{adj}\, AB = (AB)^{-1} = B^{-1}A^{-1} = \frac{1}{\det B}\, \text{adj}\, B \;\cdot\; \frac{1}{\det A}\, \text{adj}\, A = \frac{1}{\det A \cdot \det B}\, \text{adj}\, B \cdot \text{adj}\, A \tag{53}$$

As in §10.1, this suggests that it might generally be true that $\det AB = \det A \cdot \det B$ and $\operatorname{adj} AB = \operatorname{adj} B \cdot \operatorname{adj} A$. We give a proof of the first of these identities below but only a partial proof of the second is given for the special case where AB is invertible. A proof that covers the second identity in all cases can be furnished using the Cauchy-Binet formula (§13 below). In the 3×3 case we had the advantage of the formula of Theorem 17 which we obtained from the vector triple product, which allowed the AB singular case to be covered, but we have no counterpart of that for $n \times n$.

Theorem 40. Given $A, B \in \mathbb{F}^{n\times n}$, AB is invertible iff A and B are invertible.

Proof. The $\Leftarrow$ direction is clear from inverses of matrix products in $\mathbb{F}^{n\times n}$ generally (§6).

$\Rightarrow$. Suppose A singular and let $C = AB$. Then $A \sim R$ for some RREF R which has at least one zero row. Then $A = ER$ where E is a product of elementary matrices (hence E is invertible). Then $C = ERB \Rightarrow E^{-1}C = RB \Rightarrow E^{-1}C$ is singular, since R has a zero row. Thus C must be singular. Thus AB is singular if A is singular (irrespective of the status of B). Suppose now B is singular. Then B^T is singular (§6), thus by what we have just proved $B^T A^T$ is singular, ie. $(AB)^T$ is singular, and so AB is singular. Thus if A is singular or if B is singular then AB is singular and so by the contrapositive '$\Rightarrow$' is proved.

QED

Corollary. Given $A, B \in \mathbb{F}^{n\times n}$, if B is a one-sided inverse of A then A is invertible with inverse B.

Proof. Identical argument as Corollary to Theorem 21.

QED

Theorem 41. Given $A, B \in \mathbb{F}^{n\times n}$, $\det AB = \det A \cdot \det B$.

Proof. First of all from Theorem 37(c) we deduce that for any EM E and any $A \in \mathbb{F}^{n\times n}$, we have $\det EA = \det E \cdot \det A$.

Consider the case of A invertible. From Theorem 36, A is then a product of EM's, $A = E_r \cdots E_1$ say. Then :

$$\begin{aligned} & \det A &=& \det E_r \cdots \det E_1 \\ \text{and} \quad & \det AB &=& \det E_r \cdots E_1 B \\ & &=& \det E_r \cdots \det E_1 \cdot \det B = \det A \cdot \det B. \end{aligned}$$

as required.

For the case where A is singular, $\det A = 0$, and by Theorem 40, AB must be singular, thus $\det AB = 0 = \det A \cdot \det B$, as required.

QED

Corollary. Given $A, B \in \mathbb{F}^{n\times n}$ with AB invertible, $\operatorname{adj} AB = \operatorname{adj} B \cdot \operatorname{adj} A$.

Proof. Since $\det AB \neq 0$ we can multiply the equation (53) by $\det AB$ to give the desired result.

QED

13 Cauchy–Binet Formula

The proof of the Cauchy-Binet Formula [35] given below is based on the Mathematics StackExchange answers [40] and [42].

We first define some of the notations used and their properties.

Referring to the notations defined in §1 we have :

- $\left|\binom{[n]}{k}\right| = {}^nC_k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$
- $|P_k| = k!$
- $|P_{k,n}| = {}^nP_k = \frac{n!}{(n-k)!}$
- $|Q_k| = k^k$
- $|Q_{k,n}| = n^k$

- $|R_{k,n}| = {}^nC_k = \left|\binom{[n]}{k}\right|$

There is a one-to-one correspondence between the set $R_{k,n}$ and the set $\binom{[n]}{k}$ since each strictly increasing injection : $[k] \to [n]$ defines and is defined by a unique k-element subset of $[n]$. The set $P_{k,n}$ includes $R_{k,n}$ as a subset but is $k!$ times larger since for every strictly increasing injection $f \in R_{k,n}$ there are $k!$ reorderings of f, thus producing $k!$ distinct injection mappings in $P_{k,n}$.

We can denote a mapping by a list notation, eg. $(i_1, \ldots, i_k) \in Q_k$ denotes a mapping taking r to $i_r \in [k]$ for each $r \in [k]$. And $(j_1, \ldots, j_k) \in Q_{k,n}$ denotes a mapping taking r to $j_r \in [n]$ for each $r \in [k]$.

The summation notation $\sum_{i_1,i_2,\ldots,i_k=1}^{n}$ which denotes the iterated ordered sum $\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_k=1}^{n}$ can also be written as the unordered associative/commutative sum $\sum_{(i_1,i_2,\ldots,i_k)\in Q_{k,n}}$, since the order of adding up the summands and the bracketing does not matter. And because the i_r independently take on all the values from 1 to n, the mappings $(i_1, \ldots, i_k)$ are not in general injective, so that they range over $Q_{k,n}$ rather than $P_{k,n}$.

By contrast the summation $\sum_{j_1,j_2,\ldots,j_k\in\{1,2,\ldots,n\} \text{ and all distinct}}$ can be written as the associative/commutative sum $\sum_{(j_1,j_2,\ldots,j_k)\in P_{k,n}}$ with $P_{k,n}$ as the indexing set, since we are now confined to mappings $(j_1, j_2, \ldots, j_k)$ from $[k]$ to $[n]$ which are injective.

Theorem 42 (Cauchy-Binet Formula). Suppose $A \in \mathbb{F}^{k\times n}$ and $B \in \mathbb{F}^{n\times k}$, where $\mathbb{F}$ is a general field. Then :

$$\det AB = \sum_{S\in\binom{[n]}{k}} \det A_S \cdot \det B_S \tag{54}$$

where A_S denotes the $k \times k$ matrix obtained from A by selecting the column numbers in A (in order) from the set S, and B_S denotes the $k \times k$ matrix obtained from B by selecting the row numbers in B (in order) from the set S.

Proof. The case $k = n$ of the Cauchy-Binet formula reduces to the product formula for determinants (Theorem 41), ie. $\det AB = \det A \cdot \det B$, since $\binom{[n]}{n}$ just consists of the single set $[n]$.

The case $k > n$ implies $\binom{[n]}{k}$ is the empty set so the summation in (54) is empty, and therefore by convention has value zero. To show that $\det AB$ is also zero in this case, convert A to row echelon form $R \in \mathbb{F}^{k\times n}$ (§12.5) via a sequence of elementary row operations. Then there exists a product $E = E_r \cdots E_1$ of elementary matrices (each an invertible matrix in $\mathbb{F}^{k\times k}$) such that $EA = R$. Since $k > n$ and R is in row echelon form, R has at least $k-n$ zero rows at the bottom. Then $E(AB) = RB$, so if AB had an inverse $C \in \mathbb{F}^{k\times k}$ we would obtain $E = R(BC^{-1})$. But this implies E has at least $k - n$ zero rows at the bottom, which is not possible since E is invertible. Hence AB cannot have an inverse, so that $\det AB = 0$.

Consider then the case of $k < n$.

For any $J = (j_1, j_2, \ldots, j_k) \in Q_{k,n}$, we define square matrix $A(J) \in \mathbb{F}^{k\times k}$ as having r^{th} column equal to column j_r of A, $(r = 1, 2, \ldots, k)$, and square matrix $B(J) \in \mathbb{F}^{k\times k}$ as having r^{th} row equal to row j_r of B, $(r = 1, 2, \ldots, k)$. This means the columns of $A(J)$ are picked from the columns of A, but not necessarily in the order they appear in A, and a column may be picked more than once, producing an $A(J)$ with $\det A(J) = 0$. Similarly the rows of $B(J)$ are picked from the rows of B, but not necessarily in the order they appear in B, and a row may be picked more than once, producing a $B(J)$ with $\det B(J) = 0$.

Using the multilinearity of the determinant (§12.2) we first establish the following identity, using the method in [40] :

$$\det AB \quad = \quad \sum_{(j_1,j_2,\ldots,j_k)\in Q_{k,n}} (b_{j_1,1}b_{j_2,2}\cdots b_{j_k,k}) \cdot \det(A(J)) \tag{55}$$

Denoting by $(AB)_r$ the r^{th} column of $AB \in \mathbb{F}^{k\times k}$, $(r = 1, \ldots, k)$, and the standard basis vectors in $\mathbb{F}^k$ by $\underline{\mathbf{e}}_1, \ldots, \underline{\mathbf{e}}_k$, we

have, by repeatedly applying the multilinearity property of determinant :

$$\begin{aligned}
\det AB &= \det((AB)_1, (AB)_2, \ldots, (AB)_k)\,, \\
&\quad \text{using the column-wise multilinear determinant function} \\
&= \det\left(\sum_{i_1=1}^{k}\left(\sum_{j_1=1}^{n} a_{i_1,j_1}b_{j_1,1}\right)\underline{\mathbf{e}}_{i_1}, \sum_{i_2=1}^{k}\left(\sum_{j_2=1}^{n} a_{i_2,j_2}b_{j_2,2}\right)\underline{\mathbf{e}}_{i_2}, \ldots, \sum_{i_k=1}^{k}\left(\sum_{j_k=1}^{n} a_{i_k,j_k}b_{j_k,k}\right)\underline{\mathbf{e}}_{i_k}\right) \\
&= \sum_{(i_1,\ldots,i_k)\in Q_k} \det\left(\left(\sum_{j_1=1}^{n} a_{i_1,j_1}b_{j_1,1}\right)\underline{\mathbf{e}}_{i_1}, \left(\sum_{j_2=1}^{n} a_{i_2,j_2}b_{j_2,2}\right)\underline{\mathbf{e}}_{i_2}, \ldots, \left(\sum_{j_k=1}^{n} a_{i_k,j_k}b_{j_k,k}\right)\underline{\mathbf{e}}_{i_k}\right)\,, \\
&\quad \text{using the additivity of the column-wise det function in each of its } k \text{ variables} \\
&= \sum_{(i_1,\ldots,i_k)\in Q_k}\ \sum_{(j_1,\ldots,j_k)\in Q_{k,n}} \det\left(a_{i_1,j_1}b_{j_1,1}\cdot\underline{\mathbf{e}}_{i_1},\ a_{i_2,j_2}b_{j_2,2}\cdot\underline{\mathbf{e}}_{i_2}, \ldots, a_{i_k,j_k}b_{j_k,k}\cdot\underline{\mathbf{e}}_{i_k}\right)\,, \\
&\quad \text{again using the additivity of the column-wise det function in each of its } k \text{ variables} \\
&= \sum_{(i_1,\ldots,i_k)\in Q_k}\ \sum_{(j_1,\ldots,j_k)\in Q_{k,n}} (b_{j_1,1}\cdot b_{j_2,2}\cdots b_{j_k,k})\cdot\det\left(a_{i_1,j_1}\cdot\underline{\mathbf{e}}_{i_1},\ a_{i_2,j_2}\cdot\underline{\mathbf{e}}_{i_2}, \ldots, a_{i_k,j_k}\cdot\underline{\mathbf{e}}_{i_k}\right)\,, \\
&\quad \text{using the multiplicative property of the column-wise det function in each of its } k \text{ variables} \\
&= \sum_{(j_1,\ldots,j_k)\in Q_{k,n}}\ \sum_{(i_1,\ldots,i_k)\in Q_k} (b_{j_1,1}\cdot b_{j_2,2}\cdots b_{j_k,k})\cdot\det\left(a_{i_1,j_1}\cdot\underline{\mathbf{e}}_{i_1},\ a_{i_2,j_2}\cdot\underline{\mathbf{e}}_{i_2}, \ldots, a_{i_k,j_k}\cdot\underline{\mathbf{e}}_{i_k}\right)\,, \\
&\quad \text{interchanging the order of the summations} \\
&= \sum_{(j_1,\ldots,j_k)\in Q_{k,n}} (b_{j_1,1}\cdot b_{j_2,2}\cdots b_{j_k,k}) \sum_{(i_1,\ldots,i_k)\in Q_k} \det\left(a_{i_1,j_1}\cdot\underline{\mathbf{e}}_{i_1},\ a_{i_2,j_2}\cdot\underline{\mathbf{e}}_{i_2}, \ldots, a_{i_k,j_k}\cdot\underline{\mathbf{e}}_{i_k}\right)\,, \\
&= \sum_{(j_1,\ldots,j_k)\in Q_{k,n}} (b_{j_1,1}\cdot b_{j_2,2}\cdots b_{j_k,k})\cdot\det\left(\sum_{i_1=1}^{k} a_{i_1,j_1}\cdot\underline{\mathbf{e}}_{i_1},\ \sum_{i_2=1}^{k} a_{i_2,j_2}\cdot\underline{\mathbf{e}}_{i_2}, \ldots, \sum_{i_k=1}^{k} a_{i_k,j_k}\cdot\underline{\mathbf{e}}_{i_k}\right)\,, \\
&\quad \text{applying the additivity of the column-wise det function in each of its } k \text{ variables in reverse} \\
&= \sum_{(j_1,\ldots,j_k)\in Q_{k,n}} (b_{j_1,1}\cdot b_{j_2,2}\cdots b_{j_k,k})\cdot\det\left(A_{j_1}\,,\ A_{j_2}, \ldots, A_{j_k}\right)\,, \\
&\quad \text{where } A_r \text{ denotes the } r^{\text{th}} \text{ column of } A,\ (r = 1, \ldots, n) \\
&= \sum_{(j_1,\ldots,j_k)\in Q_{k,n}} (b_{j_1,1}\cdot b_{j_2,2}\cdots b_{j_k,k})\cdot\det(A(J)) \\
&\quad \text{from the definition of } A(J) \text{ above, where } J = (j_1, \ldots, j_k) \in Q_{k,n}
\end{aligned}$$

ie. the RHS of (55), as required.

From (55) we then obtain :

$$\det AB = \sum_{(j_1,j_2,\ldots,j_k)\in P_{k,n}} (b_{j_1,1}b_{j_2,2}\cdots b_{j_k,k})\cdot\det(A(J)) \tag{56}$$

since for non-injective mappings $(j_1, \ldots, j_k) \in Q_{k,n}$ we have $\det A(J) = 0$, since $A(J)$ has a duplicate column.

The remainder of the proof uses the method of [42].

Define a mapping $\psi : P_{k,n} \to R_{k,n}$ by $J = (j_1, j_2, \ldots, j_k) \mapsto \psi(J) = J' = (j'_1, j'_2, \ldots, j'_k)$, where $(j'_1, j'_2, \ldots, j'_k)$ is the ordered version of $(j_1, j_2, \ldots, j_k)$. For each $J \in P_{k,n}$, let $\sigma_J \in P_k$ be the permutation of k objects that maps the ordered version $(j'_1, j'_2, \ldots, j'_k)$ back to $(j_1, j_2, \ldots, j_k)$, ie. using the Cauchy two-line notation, [43, §2] :

$$\sigma_J = \begin{pmatrix} j'_1 & j'_2 & \cdots & j'_k \\ j_1 & j_2 & \cdots & j_k \end{pmatrix}$$

ie. as a function, σ_J maps the original position of an object in the top line of this notation to its new position in the bottom line, so that :

$$j_{\sigma_J(r)} = j'_r,\ \forall r \in [k] \tag{57}$$

Note that the numbers we are permuting with σ_J are numbers in the range $\{1, 2, \ldots, n\}$, but σ_J acts on their positions which are numbers in the range $\{1, 2, \ldots, k\}$, (σ_J is a member of the set P_k of bijections on $[k]$ — recall the dual viewpoint we have of permutations as list reorderings and as functions, [43, §1], Appendix A) . Any permutation can be expressed as a composition of swap permutations (ie. transpositions) [43, §8], so suppose σ_J is a composition of t swap permutations of k objects. Then the signature of σ_J is given by $\text{sgn}(\sigma_J) = (-1)^t$, [43, §10]. If the corresponding sequence of t column swaps is applied to the column selections for $A(J')$ then the matrix $A(J)$ is produced. But since each column swap multiplies the determinant by -1, we then have :

$$\begin{aligned} \det A(J) &= (-1)^t \det A(J') \\ \text{ie.} \quad \det A(J) &= \text{sgn}(\sigma_J) \det A(J') \\ \therefore \text{ from (56)}, \quad \det AB &= \sum_{J=(j_1,j_2,\ldots,j_k)\in P_{k,n}} (b_{j_1,1} b_{j_2,2} \cdots b_{j_k,k}) \cdot \text{sgn}(\sigma_J) \cdot \det(A(J')) \end{aligned} \tag{58}$$

Now gather together the terms in the summation (58) which have a common J'. The set of all possible J' is $R_{k,n}$ and for each $J' \in R_{k,n}$ there will be $k!$ terms $J \in P_{k,n}$ that map to that same J' under ψ, ie all the injective maps which when placed in order are the same as J'. Thus the above sum (58) partitions into the double sum :

$$\det AB = \sum_{J'=(j'_1,j'_2,\ldots,j'_k)\in R_{k,n}} \left(\sum_{\substack{J=(j_1,j_2,\ldots,j_k)\in P_{k,n}, \\ \psi(J)=J'}} (b_{j_1,1} b_{j_2,2} \cdots b_{j_k,k}) \cdot \text{sgn}(\sigma_J) \right) \cdot \det(A(J')) \tag{59}$$

Writing $j_r = j_{\sigma_J(\sigma_J^{-1}(r))}$, for $r \in [k]$, from (57) we have $j_r = j'_{\sigma_J^{-1}(r)}$, and so as the inverse of a permutation has the same signature as the permutation itself [43, §10], the inner sum of (59) becomes:

$$\sum_{\substack{J=(j_1,j_2,\ldots,j_k)\in P_{k,n}, \\ \psi(J)=J'}} (b_{j'_{\sigma_J^{-1}(1)},1} b_{j'_{\sigma_J^{-1}(2)},2} \cdots b_{j'_{\sigma_J^{-1}(k)},k}) \cdot \text{sgn}(\sigma_J^{-1}) \qquad \text{(inner sum)} \tag{60}$$

and this ranges over all $J \in P_{k,n}$ satisfying $\psi(J) = J'$. But such a range of J corresponds with all $k!$ permutations of the in-order J' (which is fixed in the inner sum), and thus σ_J over this range of J covers all $k!$ elements of the set P_k of permutations of $[k]$. Hence σ_J^{-1} over this range of J covers the set P_k of permutations also. Thus with the substitution $\tau = \sigma_J^{-1}$ the inner sum (60) becomes (for a given fixed J') :

$$\sum_{\tau\in P_k} (b_{j'_{\tau(1)},1} b_{j'_{\tau(2)},2} \cdots b_{j'_{\tau(k)},k}) \cdot \text{sgn}(\tau) \qquad \text{(inner sum)} \tag{61}$$

But from the definition of $B(J')$, $B(J')_{s,r} = b_{j'_s,r} \forall\, s, r$, and so $B(J')_{\tau(r),r} = b_{j'_{\tau(r)},r}$ for $r \in [k]$, and thus the inner sum (61) equals :

$$\sum_{\tau\in P_k} (B(J')_{\tau(1),1} B(J')_{\tau(2),2} \cdots B(J')_{\tau(k),k}) \cdot \text{sgn}(\tau) \qquad \text{(inner sum)}$$

which is the column-wise Leibniz formula for the $k \times k$ determinant $\det B(J')$. Thus from (59) :

$$\begin{aligned} \det AB &= \sum_{J'\in R_{k,n}} \det B(J') \cdot \det A(J') \\ &= \sum_{S\in\binom{[n]}{k}} \det A_S \cdot \det B_S \end{aligned}$$

using the one-to-one correspondence between sets $R_{k,n}$ and $\binom{[n]}{k}$, where A_S denotes the $k \times k$ matrix obtained from A by selecting the column numbers in A (in order) from the set S, and B_S denotes the $k \times k$ matrix obtained from B by selecting the row numbers in B (in order) from the set S. Thus $\det AB$ equals the required formula (54).

QED

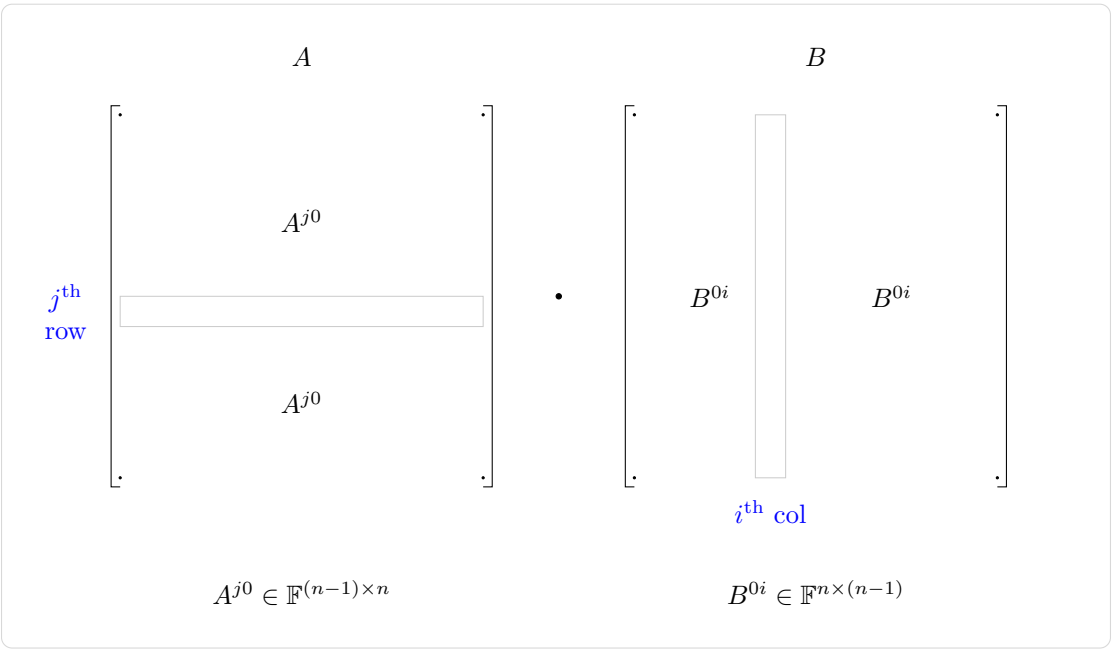
13.1 Proof Of Adjugate Product Formula

In §12.7 we were able to prove the adjugate product formula $\text{adj}\, AB = \text{adj}\, B \cdot \text{adj}\, A$ in the special case where AB was invertible. The following proof uses the Cauchy-Binet formula and covers the general case. The method of the proof is as described in the Mathematics StackExchange answer [41].

Theorem 43. Given $A, B \in \mathbb{F}^{n\times n}$, $\operatorname{adj} AB = \operatorname{adj} B \cdot \operatorname{adj} A$.

Proof.

Figure 18: $(j,i)^{\text{th}}$ Minor Matrix Of $AB = (AB)^{ji} = A^{j0} \cdot B^{0i} \in \mathbb{F}^{(n-1)\times(n-1)}$



In the following denote by the superscript notation[2] A^{ij} the $(i,j)^{\text{th}}$ minor matrix of A, ie. the submatrix of A obtained by removing row i and column j. Allow i and j to take on the value zero, so that A^{i0} is obtained by removing only row i but none of the columns, and A^{0j} is obtained by removing only column j but none of the rows. From Figure 18 we then see that $(AB)^{ji} = A^{j0} \cdot B^{0i}$. We will use the subscript notation A_{ij} to denote the $(i,j)^{\text{th}}$ element of A.

Then :

$$\begin{aligned}
\det(AB)^{ji} &= \det(A^{j0} \cdot B^{0i}) \\
&= \sum_{S\in\binom{[n]}{n-1}} \det A^{j0}_S \cdot \det B^{0i}_S\,, \quad \text{by the Cauchy-Binet formula} \\
&= \sum_{k=1}^{n} \det A^{jk} \cdot \det B^{ki}\,, \qquad (62) \\
&\quad \text{where } k \text{ denotes the single element of } [n] \text{ which defines (by its omission)} \\
&\quad \text{an } (n-1) \text{ element subset of } [n]
\end{aligned}$$

[2]Note this is a different notation for minor matrix than we used in §11 and §12.3.

And we have :

$$\begin{aligned}
(\operatorname{adj} B \cdot \operatorname{adj} A)_{ij} &= (\operatorname{adj} B \cdot \operatorname{adj} A)^T_{ji} \\
&= ((\operatorname{adj} A)^T \cdot (\operatorname{adj} B)^T)_{ji} \\
&= (\mathrm{CF}(A) \cdot \mathrm{CF}(B))_{ji}\,, \quad \text{from Definition 19} \\
&= \sum_{k=1}^{n} \mathrm{CF}(A)_{jk} \cdot \mathrm{CF}(B)_{ki} \\
&= \sum_{k=1}^{n} (-1)^{j+k} \det A^{jk} \cdot (-1)^{k+i} \det B^{ki} \\
&= (-1)^{i+j} \sum_{k=1}^{n} \det A^{jk} \cdot \det B^{ki} \\
&= (-1)^{j+i} \det(AB)^{ji}\,, \quad \text{from (62)} \\
&= (j,i)^{\text{th}} \text{ of } \mathrm{CF}(AB) \\
&= (i,j)^{\text{th}} \text{ of } \operatorname{adj}(AB)
\end{aligned}$$

and so $\operatorname{adj} AB = \operatorname{adj} B \cdot \operatorname{adj} A$ as required.

QED

Appendix A

Permutations

In this article we use position-based permutations (PBP's) as described in [43]. We think of a permutation as being a native object which is a reordering of a list of n objects, and which is 'agnostic' to the actual nature of the objects themselves and only affects their positions in the list. The word 'permutation' is used in a quite different sense in many texts to mean a bijective mapping on a general set X, but this is a quite different meaning to the one used here where we are operating on a sequence rather than a set.

We denote the set of all permutations of n objects by PERM(n), a set of size $n!$, [43, §1]. With each $P \in \text{PERM}(n)$ there is associated a unique permutation function $\sigma \in S_n$, where S_n is the set of all bijections on the set $[n]$, which maps the original position r of an object in the list to the new position $\sigma(r)$ after the permutation is applied. This gives a one-to-one mapping from PERM(n) to S_n which is a group isomorphism, PERM(n) forming a group under composition of permutations and S_n forming a group under composition of functions, [43, §5]. Because of this one-to-one correspondence we often refer to 'permutation' and 'permutation function' interchangeably, ie. sometimes we view the permutation as a reordering of a list and sometimes as a function on a set.

We use a two-line notation to specify a permutation, such as :

$$P = \begin{pmatrix} 1 & 2 & \cdots & n \\ a_1 & a_2 & \cdots & a_n \end{pmatrix},$$

where $a_r \in [n]\ \forall\ r \in [n]$ and the a_r's are all distinct. The first line of the notation shows the original positions of the objects in the list and the second line shows their positions after the permutation P has been applied. For this particular notation we would have $\sigma(a_r) = r,\ \forall\ r \in [n]$, since a_r appeared at position a_r in the original list in the top line and moved to position r in the second line.

Since the two-line notation is just a way of specifying the rearrangement that is performed by the permutation we can use it with any kind of objects at all, for example $\begin{pmatrix} \text{dog} & \text{sparrow} & \text{cat} \\ \text{cat} & \text{dog} & \text{sparrow} \end{pmatrix}$ and $\begin{pmatrix} \underline{\mathbf{u}} & \underline{\mathbf{v}} & \underline{\mathbf{w}} \\ \underline{\mathbf{w}} & \underline{\mathbf{u}} & \underline{\mathbf{v}} \end{pmatrix}$ denote the same permutation of 3 objects as $\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$, with the permutation function σ given by $\sigma(1) = 2, \sigma(2) = 3, \sigma(3) = 1$. This allows suitable notation to be chosen that matches the problem at hand, eg. as we used for the permutation σ_J in the proof of the Cauchy-Binet formula in §13. We do not need to always have the top line a consecutive list of integers of form $\begin{pmatrix} 1 & 2 & \cdots & n \end{pmatrix}$, we could have any list of integers, or a list of any object types at all so long as they are all distinct so that the permutation is uniquely specified.

Some sources use value-based permutations (VBP's) rather than position-based permutations and their permutation functions refer to the actual values in the list rather than their positions. This causes confusion when there are duplicate objects in the list. Also the value-based approach produces a group 'anti-isomorphism' between PERM(n) and S_n, so that S_n becomes the 'anti-group' of PERM(n). The anti-group of a group is the same group but with the group operation reversed. This results in confusion in identifying permutations with their permutation functions since $P_m \cdots P_1$ is identified with $\sigma_1 \cdots \sigma_m$, resulting in $\sigma_m \cdots \sigma_1$ being identified with $\sigma_1 \cdots \sigma_m$, in which these are in the opposite order, which is a problem since permutations do not commute. All the permutations in this article follow the position-based approach. The article [43] contrasts PBP's with VBP's and investigates the relationship between them (particularly see §3 and §5).

Any permutation σ can always be expressed as a composition (or 'product') of transpositions, which are permutations which swap two of the objects in the list but leave all the other objects in the same position. It can be shown that the number of transpositions in such a product always has the same parity (ie. even or odd) for a given permutation σ, [43, Theorem 3, §9], and this is defined to be the parity of σ. The signature $\text{sgn}(\sigma)$ of σ is defined to be -1 for odd parity and $+1$ for even parity. The signature function then has the properties $\text{sgn}(\sigma\tau) = \text{sgn}(\sigma) \cdot \text{sgn}(\tau)$ and $\text{sgn}(\sigma^{-1}) = \text{sgn}(\sigma)$.

Appendix B

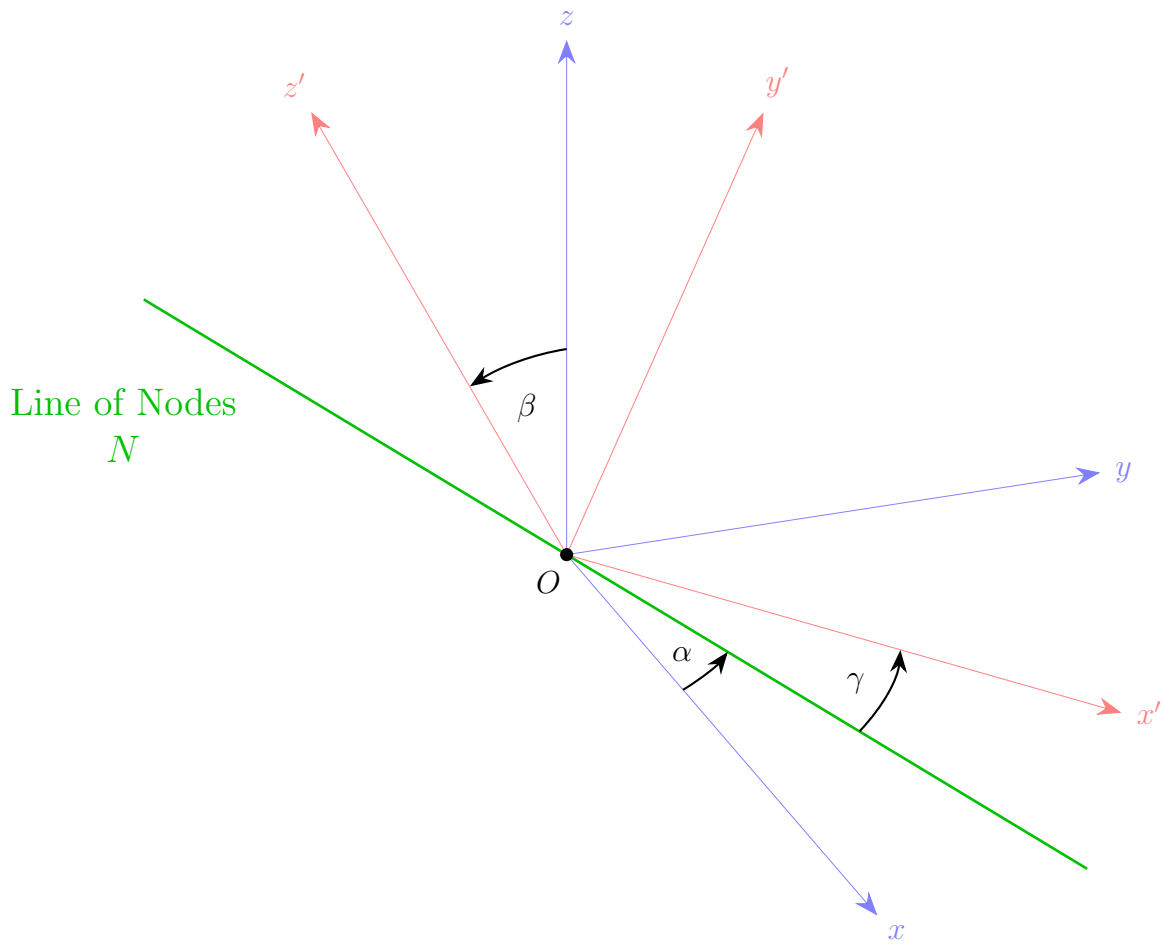
Euler Angles

A CCS can be rotated three-dimensionally in ES about its origin in an arbitrary manner to produce another CCS with the same handedness. Such a rotation can always be achieved via a sequence of three axial rotations about angles known as Euler angles α, β, γ, as shown in Figure 19 with a RH CCS. The original CCS has axes x, y, z and the rotated CCS$'$ has axes x', y', z'. We define the line of nodes N to be the intersection of the xy-plane and the $x'y'$-plane, in the case where these planes are not coincident (we consider the case where they are coincident separately below). The sequence of three axial rotations that will transform CCS into CCS$'$ is as follows (using the RH rule convention for positive angle direction for rotation about a directed axis, §3.3) :

1. Rotate the xy-plane about the z-axis by angle α so that the x-axis coincides with N
2. Rotate the xy-plane about the new x-axis by angle β so that the positive z-axis now coincides with the positive z'-axis
3. Rotate the new xy-plane about the z'-axis by angle γ so that the positive x-axis now coincides with the positive x'-axis

Then the positive y-axis will be constrained to align with the positive y'-axis, so that CCS is now fully coincident with CCS$'$. This rotation sequence is called a zxz-rotation.

Figure 19: Euler Angles α, β, γ



B.1 Uniqueness Of Euler Angles

The angles α, β, γ in the range $[0°, 360°)$ are not unique, there is one further set of angles δ, ϵ, ζ in $[0°, 360°)$ that will also take CCS to CCS$'$ via a zxz-rotation. Firstly δ must be either be α or $\alpha + 180°$ (mod 360°) for otherwise the subsequent rotation β would not be rotating about axis N as is required to reach z'. If $\delta = \alpha$ then we are rotating about the same directed axis as before so ϵ must equal β to reach z' and then finally we must also have $\zeta = \gamma$ for x to end up

in the final correct position of x'. If $\delta = \alpha + 180^\circ$ then we are rotating by ϵ about N but with the direction of axis N reversed. Thus we must have $\epsilon = 360^\circ - \beta$ to reach z'. And finally because in the first z rotation we placed the positive x-axis 180° from where it was before we must have $\zeta = \gamma + 180^\circ$ to place x correctly onto x'. Thus the second set of Euler angles in the range $[0^\circ, 360^\circ)$ which takes CCS to CCS$'$ is :

$$(\delta, \epsilon, \zeta) \quad = \quad (\alpha + 180^\circ, 360^\circ - \beta, \gamma + 180^\circ) \pmod{360^\circ} \tag{63}$$

Since we have excluded the case where the xy-plane and the $x'y'$-plane coincide, the z and z' axes cannot be collinear, so $\beta \in (0^\circ, 360^\circ) \setminus \{180^\circ\}$, and so $\beta \in (0^\circ, 180^\circ)$ or $360^\circ - \beta \in (0^\circ, 180^\circ)$. Thus there is a unique set of Euler angles (α, β, γ) with $\alpha, \gamma \in [0^\circ, 360^\circ)$ and $\beta \in (0^\circ, 180^\circ)$. The second set of possible angles $(\delta,\ \epsilon,\ \zeta)$ from (63) has $\delta, \zeta \in [0^\circ, 360^\circ)$ and $\epsilon \in (180^\circ, 360^\circ)$.

When the xy-plane and the $x'y'$-plane coincide we must have $\beta = 0^\circ$ or $\beta = 180^\circ$. In the case where $\beta = 0^\circ$ the z-axis does not move, and under the three rotation steps above the x-axis rotates about z by α and then about z by γ to reach x'. Thus if the direction of x' from x is an angle Ω wrt rotation about the original positive z-axis then the set of all possible Euler angles in the range $[0^\circ, 360^\circ)$ which moves CCS to CCS$'$ is:

$$\{(\alpha, 0^\circ, \gamma) : \alpha, \gamma \in [0^\circ, 360^\circ), \alpha + \gamma = \Omega \pmod{360^\circ}\}.$$

In the case where $\beta = 180^\circ$ the z-axis flips direction 180° between the first rotation of α and the third rotation of γ. Thus the third rotation is $-\gamma$ when taken relative to the original positive z-axis. Thus the combination of the three rotation steps above is the first and third rotations with a z-axis flip in between, which is equivalent to a single rotation about the original positive z-axis of angle $\alpha - \gamma$. Thus the set of all possible Euler angles in the range $[0^\circ, 360^\circ)$ which moves CCS to CCS$'$ is:

$$\{(\alpha, 180^\circ, \gamma) : \alpha, \gamma \in [0^\circ, 360^\circ), \alpha - \gamma = \Omega \pmod{360^\circ}\}.$$

In engineering applications of 3D rotations and Euler angles, such as gyroscopes, compasses, and robotics the condition corresponding to the xy-plane and the $x'y'$-plane coinciding is called 'gimbal lock' [36].

B.2 Elementary Rotation Matrices

The three rotations in the above zxz-rotation can each be expressed in terms of a 2D coordinate transform matrix of form $\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$ which converts (x, y) coordinates into (x', y') coordinates wrt $x'y'$-axes which are rotated about the origin from the xy-axes by angle θ, ie.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} \quad = \quad \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

In the first rotation by α the z-axis is held fixed so the transform matrix producing the first set of intermediate coordinates is :

$$C \quad = \quad \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

In the second rotation by β the x-axis is held fixed so the transform matrix producing the second set of intermediate coordinates is :

$$B \quad = \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\beta & \sin\beta \\ 0 & -\sin\beta & \cos\beta \end{bmatrix}$$

And in the third rotation by γ the z-axis is held fixed so the transform matrix producing the final set of coordinates wrt CCS$'$ is :

$$A \quad = \quad \begin{bmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The transform matrix M which converts coordinates $\underline{\mathbf{s}} = (x, y, z) \in \mathbb{R}^3$ of $\underline{\mathbf{r}} \in$ EV wrt CCS to coordinates $\underline{\mathbf{s}}' = M\,\underline{\mathbf{s}} = (x', y', z') \in \mathbb{R}^3$ wrt CCS$'$ is then given by $M = ABC$. Thus from §4.5.1 any RH orthogonal matrix in $\mathbb{R}^{3\times3}$ can be expressed as such a product ABC of three elementary rotation matrices.

The Euler angle method described above would work equally well for both RH and LH CCS. For LH CCS we could continue to use the standard RH convention for positive angle sense so that the above elementary rotation matrices A, B, C would work as before.

Appendix C

Completeness Of Real Numbers

This appendix is provided for background interest in relation to the geometric discussion of §2 but is not required for the main article.

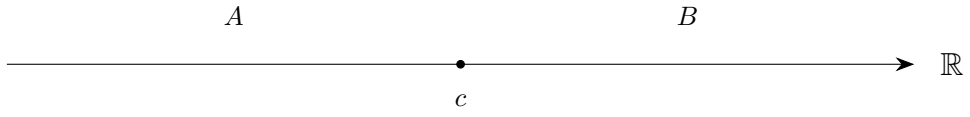
The set $\mathbb{R}$ is defined as the set of all points on a one-dimensional infinite straight line within ES (ie. ES_1, the one dimensional Euclidean space) which has had a direction assigned to it denoted as 'positive'. A total order relation $\leq$ [14] is defined wrt to the positive direction. The theorem below applies prior to any selection of unit of distance on the line or definition of arithmetic operations on $\mathbb{R}$ or the assignment of a zero point on the line.

The proof relies upon the following postulate which is a reasonable assumption intuitively due to the fact that the real number line by definition has no gaps.

Postulate (Boundary Property Of $\mathbb{R}$). If the real number line $\mathbb{R}$ is partitioned into two non-empty sets A and B with A strictly to the left of B (ie. $a \in A$ and $b \in B \Rightarrow a < b$) then there exists a 'crossover point' $c \in \mathbb{R}$ such that $x < c \Rightarrow x \in A$ and $x > c \Rightarrow x \in B$.

□

Figure 20: Crossover Point c



From an intuitive perspective, with the set A strictly to the left of set B there is no possibility of A spilling over into B nor of B spilling over into A so that there must be a sharp boundary of some kind in passing from A to B, and this can reasonably only take the form of a single point.

The sets A and B above have the property of being 'continuous to the left' and 'continuous to the right' respectively however the set S in Theorem C.1 below can be an arbitrarily complicated set. Intuitively, the least upper bound of S in Theorem C.1 below is a point located exactly at the rightmost extent of S on the real number line, and we can imagine a slider on the real number line, as on a slide rule, that demarcates this point.

Theorem C.1 (Completeness Property Of $\mathbb{R}$). Every non-empty $S \subseteq \mathbb{R}$ which is bounded above has a least upper bound.

Proof. Let :

$$\begin{aligned} A &= \{x \in \mathbb{R} : x \text{ is not an upper bound of } S\} \\ B &= \{x \in \mathbb{R} : x \text{ is an upper bound of } S\} \end{aligned}$$

Then A and B are disjoint and non-empty and form a partition of $\mathbb{R}$. If $a \in A$ and $b \in B$, $\exists\ s \in S$ with $a < s \leq b$, and hence $a < b$. Thus A is strictly to the left of B. Thus there must exist a crossover point $c \in \mathbb{R}$ as in the postulate above.

If $c \in A$ then c is not an upper bound for S, so $\exists\ s \in S$ such that $c < s$. Then $\exists\ d \in \mathbb{R}$ such that $c < d < s$. Then $d \in B$, but d being less than s cannot be an upper bound for S — a contradiction. Thus $c \notin A$ so we must have $c \in B$, ie. c is an upper bound for S. Furthermore if $e < c$ then $e \in A$, ie. e is not an upper bound for S. Hence c is the least upper bound for S.

QED

A proof along similar lines could be constructed to establish the completeness property for the set of all angles $\mathbb{A}$ (§2).

It can be readily shown that the converse of Theorem C.1 also holds in the sense that if we assume completeness of $\mathbb{R}$ then the boundary property of $\mathbb{R}$ defined in the above postulate follows. Thus for $\mathbb{R}$ the completeness and boundary

properties are equivalent to one another. This fact applies in any totally ordered set $(X, \leq)$, [14], not just in $\mathbb{R}$, and the proof above carries over unchanged to this more general environment — notice though how the totality property of $\leq$ is required for the proof to work, ie. that any two elements of the set X are comparable, so that the law of trichotomy [25] holds. This equivalence on $(X, \leq)$ is not dependent on geometry in any way and it is a purely 'a priori' deduction in this more formal setting.

References

All web links retrieved on 22nd June 2026.

[1] MacTutor archive, University of St. Andrews, Scotland, *Wilhelm Ahrens book of quotes*, https://mathshistory.st-andrews.ac.uk/Extras/Ahrens_quotes/

[2] Paul R. Halmos (1960), *Naive Set Theory*, Van Nostrand Reinhold Company.

[3] Thomas L. Heath (2013), *Euclid's Elements*, Green Lion Press, Santa Fe, New Mexico.

[4] Iain T. Adamson (1996), *A General Topology Workbook*, Birkhäuser Boston.

[5] Iain T. Adamson (1982), *Introduction To Field Theory*, 2nd Edition, Cambridge University Press.

[6] Stephen Willard (1970), *General Topology*, Addison-Wesley Publishing Company, Inc.

[7] Paul B. Yale (1968), *Geometry And Symmetry*, Holden Day.

[8] Robert G. Bartle & Donald R. Sherbert (1982), *Introduction to Real Analysis*, John Wiley & Sons.

[9] Robert G. Bartle (1964), *The Elements of Real Analysis*, John Wiley & Sons.

[10] Thomas A. Whitelaw (1994), *Introduction To Linear Algebra*, 2nd Edition, Blackie Academic & Professional

[11] Thomas A. Whitelaw (1978), *An Introduction To Abstract Algebra*, Blackie & Sons Ltd.

[12] Earl W. Swokowski (1983), *Calculus With Analytic Geometry*, Alternate Edition, Prindle, Weber, & Schmidt Ltd.

[13] *Line segment/Directed line segment*, https://en.wikipedia.org/wiki/Line_segment#Directed_line_segment

[14] *Total order*, https://en.wikipedia.org/wiki/Total_order.

[15] *Construction of the real numbers*, https://en.wikipedia.org/wiki/Construction_of_the_real_numbers.

[16] *Equipollence (geometry)*, https://en.wikipedia.org/wiki/Equipollence_(geometry).

[17] *Norm (mathematics)*, https://en.wikipedia.org/wiki/Norm_(mathematics).

[18] *Chirality*, https://en.wikipedia.org/wiki/Chirality.

[19] *Cartesian coordinate system*, https://en.wikipedia.org/wiki/Cartesian_coordinate_system.

[20] *Coordinate vector*, https://en.wikipedia.org/wiki/Coordinate_vector.

[21] *Orthogonal matrix*, https://en.wikipedia.org/wiki/Orthogonal_matrix.

[22] *Conformable matrix*, https://en.wikipedia.org/wiki/Conformable_matrix.

[23] *Finite field*, https://en.wikipedia.org/wiki/Finite_field.

[24] *Ordered field*, https://en.wikipedia.org/wiki/Ordered_field.

[25] *Law of trichotomy*, https://en.wikipedia.org/wiki/Law_of_trichotomy.

[26] *Inner product space*, https://en.wikipedia.org/wiki/Inner_product_space.

[27] *Orthogonality (mathematics)*, https://en.wikipedia.org/wiki/Orthogonality_(mathematics).

[28] *Symmetric group*, https://en.wikipedia.org/wiki/Symmetric_group

[29] *Multilinear map*, https://en.wikipedia.org/wiki/Multilinear_map

[30] *Row echelon form*, https://en.wikipedia.org/wiki/Row_echelon_form

[31] *Elementary matrix*, https://en.wikipedia.org/wiki/Elementary_matrix

[32] *Gaussian elimination*, https://en.wikipedia.org/wiki/Gaussian_elimination

[33] *Parity of a permutation/Equivalence of the two definitions*, https://en.wikipedia.org/wiki/Parity_of_a_permutation#Equivalence_of_the_two_definitions

[34] *Parity of a permutation*, https://en.wikipedia.org/wiki/Parity_of_a_permutation

[35] *Cauchy-Binet Formula*, https://en.wikipedia.org/wiki/Cauchy-Binet_formula

[36] *Gimbal lock*, https://en.wikipedia.org/wiki/Gimbal_lock

[37] *Permutation/Two-line notation*, https://en.wikipedia.org/wiki/Permutation#Two-line_notation

[38] *M.7 Gauss-Jordan Elimination*, Department of Statistics, Eberly College of Science, Penn State University, https://online.stat.psu.edu/statprogram/reviews/matrix-algebra/gauss-jordan-elimination

[39] Gowers's Weblog : *Permutations*, https://gowers.wordpress.com/2011/10/16/permutations/

[40] W. G. (https://math.stackexchange.com/users/301204/w-g), Mathematics StackExchange, answer to *Cauchy-Binet Formula Proof Intuition (Determinants)*, https://math.stackexchange.com/a/3245450/417024.

[41] VanD1206 (https://math.stackexchange.com/users/428004/vand1206), Mathematics StackExchange, answer to *adj(AB) = adjB adjA*, https://math.stackexchange.com/a/2719486/417024.

[42] Ross Ure Anderson (https://math.stackexchange.com/users/417024/ross-ure-anderson), Mathematics StackExchange, answer to *Cauchy-Binet Formula Proof Intuition (Determinants)*, https://math.stackexchange.com/a/4014267/417024.

[43] Ross Ure Anderson (2026), *Position-Based Permutations*, https://archive.org/details/permutations_202601 and https://www.academia.edu/168691304/Position_Based_Permutations

[44] Ross Ure Anderson (https://math.stackexchange.com/users/417024/ross-ure-anderson), Mathematics StackExchange, question *Can The Existence Of π Be Proved Without Formal Analysis?*, https://math.stackexchange.com/q/2972693/417024, and answer https://math.stackexchange.com/a/3014865

[45] Ross Ure Anderson (https://math.stackexchange.com/users/417024/ross-ure-anderson), Mathematics StackExchange, answer to *How to prove that the area of a shape is independent of the choice of axes?*, https://math.stackexchange.com/a/4862275/417024.

[46] Ross Ure Anderson (https://math.stackexchange.com/users/417024/ross-ure-anderson), Mathematics StackExchange, answer to *Surface integrals in spherical coordinates*, https://math.stackexchange.com/a/4867164/417024.

[47] Ross Ure Anderson (https://math.stackexchange.com/users/417024/ross-ure-anderson), Mathematics StackExchange, answer to *Surface integrals in spherical coordinates*, https://math.stackexchange.com/a/4871658/417024.

[48] Ross Ure Anderson (2020), *Derivation of Solar Position Formulae*, https://arxiv.org/abs/2009.07094 and https://archive.org/details/derivation-solar-position-formulae